

Joint Inversion of Receiver Function and Surface Wave Dispersion by Hamiltonian Monte Carlo Sampling

Junliu Suwen^{1,2,3}, Qi-Fu Chen^{1,2,3}, and Nanqiao Du^{*2,3,4,5}

Abstract

We have proposed a new probabilistic inversion method to perform the joint inversion of receiver function and surface wave dispersion data. In this method, we apply the Hamiltonian dynamics in the Bayesian framework to efficiently sample the posterior probability distribution of this joint inverse problem. This method will lead to nearly 100% acceptance of each sample in theory. Semianalytical derivatives of both the datasets to the model parameters (including elastic parameters, density, and the thickness of each layer) are used to speed up this algorithm. Finally, we apply our method to both synthetic data and real data. The result shows that the velocity model can be recovered well within a much smaller number of samplings than the traditional Markov chain Monte Carlo method.

Cite this article as Suwen, J., Q.-F. Chen, and N. Du (2022). Joint Inversion of Receiver Function and Surface Wave Dispersion by Hamiltonian Monte Carlo Sampling, *Seismol. Res. Lett.* **XX**, 1–16, doi: [10.1785/0220220044](https://doi.org/10.1785/0220220044).

Introduction

The elastic parameters and the corresponding velocity variation in the shallow and deep earth are quite essential to our understanding of the dynamics of the earth system. Seismic imaging is one of the most powerful tools to investigate these characteristics in a data-driven approach, thereby leading to many successful achievements including travel-time tomography (Thurber and Um, 1987; Rawlinson and Sambridge, 2004, 2005), waveform inversion (Fichtner *et al.*, 2008, 2009, 2013), and attenuation analysis (Zhao *et al.*, 2013; Zhao and Mousavi, 2018).

Among these approaches, the receiver function (RF) analysis and surface wave dispersion (SWD) inversion together have become one of the most popular methods in the past decades due to the wide availability of observations, the complementarity of these two datasets, and the high sensitivity to the variation of elastic parameters. RF is a time series that contain reflection and conversion waves by deconvolving the vertical component from the horizontal component (Langston, 1979). The deconvolution removes the effect of source and travel path, and it accentuates the relative arrival time between the first arrivals and multiples. The relative arrival time is sensitive to velocity discontinuities (Ammon *et al.*, 1990) but contains fewer constraints on absolute velocity. On the contrary, the SWDs are sensitive to absolute shear-wave velocities (Shen *et al.*, 2013), and thus, be widely applied to the crustal structure of *S* waves (Bourjot and Romanowicz, 1992). Therefore, it is natural to combine these two data sets to perform the joint inversion. Joint inversion could benefit from both the methods and overcome their limitations when two datasets are inverted separately.

Traditionally, joint inversion will be cast into a linearized, iterative least square problem (Julià *et al.*, 2000). However, an initial model is required for the iterative method. The result obtained by the minimization of the misfit function is highly relevant to the choice of the initial model (Ammon *et al.*, 1990) and may not always converge to the global minimum. Another problem is the incomplete observation, which leads to the nonuniqueness of this inverse problem (Backus and Gilbert, 1967; Jackson, 1979). Although the regularization in objective function would reduce the ambiguity of solution, it may introduce unrealistic smoothing to the result (Sen *et al.*, 2014). These challenges prompt us to solve the inversion problem from another point of view—the probabilistic framework (Tarantola, 2005).

The Bayesian inference is more and more prevalent in geophysical inversion recently (Sambridge, 1999; Sambridge and Mosegaard, 2002; Bodin, Sambridge, Rawlinson, *et al.*, 2012; Shen *et al.*, 2013). It describes the connection between the posterior probability density and prior information or likelihood function. In the view of the Bayesian framework, posterior probability density represents the possibility of model parameters for given observation data, and the likelihood function is

1. Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China; 2. Institutions of Earth Science, Chinese Academy of Sciences, Beijing, China; 3. College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing, China; 4. Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China; 5. Department of Earth Sciences, University of Toronto, Toronto, Ontario, Canada

*Corresponding author: nanqiao.du@gmail.com

© Seismological Society of America

established by misfit function of the specific inversion problem. The basic idea behind Monte Carlo method is that the characteristics of posterior probability density can be obtained by sampling methods. The Markov chain Monte Carlo (McMC; Hastings, 1970) is the most commonly used sampling method. With the Metropolis–Hastings criterion (Chib and Greenberg, 1995), sampling density will converge to target distribution after massive trials of McMC approach. Moreover, the model parameterization (i.e., the number of layers) can be acquired by the transdimensional McMC method. The reversible-jump algorithm (Green, 1995) allows the change of dimensions of parameter subspace by using a reversible Markov chain and is also widely used in geophysics inversion (Bodin and Sambridge, 2009; Agostinetti and Malinverno, 2010; Bodin, Sambridge, Tkalcic, *et al.*, 2012). Nevertheless, the computational cost of sampling in the Monte Carlo method is high, and the convergence rate is extremely low, especially for sampling complex posteriors in high dimensional space. The way-out to perform the faster sample is to make use of gradient information (Girolami and Calderhead, 2011).

The Hamiltonian Monte Carlo (HMC) is one of the McMC methods and has been used in solving the seismic inverse problems for the recent several years (Sen and Biswas, 2017; Fichtner and Simute, 2018; Fichtner and Zunino, 2019; Aleardi *et al.*, 2020). It brings Hamiltonian dynamics to sample processes, which helps to reduce the random behavior and make the sampling more effective. In the HMC context, the model is treated as a particle that has mass, momentum, and potential energy. The inverse process can be considered as the particle moving from the current position to a new position in the phase space. The potential energy is interpreted as the misfit function, whereas the kinetic energy provides the updated direction of the model parameters in the sampling framework (Girolami and Calderhead, 2011). Although the cost of gradient computation may be much high, especially when using numerical differentiation to approximate the gradient, it can be suitable for the problems that have the analytic or semianalytic form of derivative.

This article is organized in the following ways: First, we introduce the HMC framework including how the forward and inverse problem is defined, how to merge HMC in our problem, and calculate the gradient of both the datasets. Then we compare our method with one of the prevalent McMC methods (reversible-jump [rj] McMC) in the synthetic test problem to show the advantage of our HMC method and apply it to the real data set. Finally, we will discuss the effects of some parameters in our framework and focus on the limitation of our method.

Methodology

HMC basics for inversion problems

In the probabilistic point of view, a geophysical inversion problem could be reckoned as a Bayesian inference problem in

which we seek to find a (or a series of) model(s) to adequately sample a target posterior probabilistic distribution (Tarantola, 1984). To be specific, in the Bayesian inference framework the posterior of the inverse problem is

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

in which \mathbf{m} denotes the unknown model parameters, \mathbf{d}_{obs} is the dataset (usually with noise) from geophysical observations, $P(\mathbf{m})$ is the prior information we have known about these unknown parameters, and $P(\mathbf{d}_{\text{obs}}|\mathbf{m})$ is the possibility of obtaining the dataset by a given model, thus, usually including the physics and error estimation of the inverse problem. It could be related to the misfit function through

$$P(\mathbf{m}|\mathbf{d}_{\text{obs}}) \propto \exp(-\Phi(\mathbf{m}, \mathbf{d}_{\text{obs}})), \quad (2)$$

in which Φ is the misfit function of the inverse problem. If Gaussian noise is preferred, then it will have the form:

$$\Phi = \frac{1}{2}(\mathbf{d} - \mathbf{d}_{\text{obs}})^T \Sigma^{-1}(\mathbf{d} - \mathbf{d}_{\text{obs}}), \quad (3)$$

in which Σ is the covariance matrix of the dataset.

Several approaches are presented to locate these “optimized models”. The most prevalent technique for model inference is maximum-likelihood estimation, which aims at finding a model to maximize the likelihood function. Linearized inversion is one of the preferred methods to find the best models by utilizing the gradient of the posteriors (Julià *et al.*, 2000; West *et al.*, 2004; Nunn *et al.*, 2014), but it may be stuck in the local minimal due to the complexity of the posterior function (Aster *et al.*, 2012). On the contrary, global optimization methods, including Nearest Algorithm (Sambridge, 1999) and transdimensional Monte Carlo (Bodin, Sambridge, Tkalcic, *et al.*, 2012), were developed and prevalent in recent years partly due to the significant improvement of computational resources, especially in 1D and 2D problems (including RF and SWD problem). However, these start-of-the-art approaches all face a problem that they are all derivative-free search algorithms and thus did not include useful derivative information of the posterior (Fichtner *et al.*, 2019). As a result, a low accept ratio or slow searching rate may occur in sampling the posterior distribution.

There are some alternative methods that use gradients such as the variational method (Nawaz and Curtis, 2018; Zhang and Curtis, 2020; Zhang *et al.*, 2021), Langevin Monte Carlo method (Siahkoohi *et al.*, 2020), and the Hamiltonian Monte Carlo method is one of those derivative-based approaches. In this method, model parameters are viewed as the n -dimensional location of a particle moving in the phase space (momentum–location space; Neal, 2012). The trajectories of this particle are determined by the Hamiltonian dynamics:

$$\frac{dm_i}{dt} = \frac{\partial K}{\partial p_i}; \frac{dp_i}{dt} = -\frac{\partial U}{\partial m_i}, \quad (4)$$

in which the Hamiltonian is the total energy of the system (kinetic energy $K(\mathbf{p})$ and potential energy $U(\mathbf{m})$), and p_i is a random variable that plays the role of momentum of the particle. In the Bayesian framework of equation (1), the kinetic and potential energy are defined as (Fichtner and Zunino, 2019):

$$U(\mathbf{m}) = -\ln P(\mathbf{m}|\mathbf{d}_{\text{obs}}); K(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T \mathbf{M}^{-1}\mathbf{p}, \quad (5)$$

in which M is the mass matrix that determines the different weights of different parameters. To search for a next model, we randomly generate a new momentum for current particle and make it evolve by the mechanics of equation (4) to a new location, and the Metropolis rule is used to accept or reject this new location with the probability:

$$\Pi_{\text{accept}} = \min\{\exp(-H_{\text{new}} + H_{\text{current}}), 1\}. \quad (6)$$

In addition, equation (4) indicates that the Hamiltonian of the particle is constant along the trajectory for a given starting point and momentum. Thus, the huge advantage of HMC over other MCMC methods is that the accept ratio of the new point is nearly 1 (In practice, numerical errors may accumulate when numerically solving this system).

There are several applications of HMC in geophysical inversions, including seismic source inversion (Fichtner and Simute, 2018), Rayleigh-wave dispersion inversion (Aleardi et al., 2020), P-wave tomography (Muir and Tkalcic, 2020), and even a full-waveform inversion (Gebraad et al., 2020). However, it still lacks the applications of HMC on the RF and SWD joint inversion problems. In the HMC for joint inversion problem and the Forward method sections, we will thoroughly discuss how HMC could be merged with this classical inverse problem.

HMC for joint inversion problem

In the joint inversion problem of RF and SWD, a commonly used misfit function is from Julià et al. (2000):

$$\Phi = \frac{1}{2}\sqrt{\frac{p}{w_r^2}}\sum_{i=1}^m (\mathbf{d}_i^r - \mathbf{d}_i^{r,\text{obs}})^2 + \frac{1}{2}\sqrt{\frac{1-p}{w_s^2}}\sum_{i=1}^m (\mathbf{d}_i^s - \mathbf{d}_i^{s,\text{obs}})^2, \quad (7)$$

in which w_r, w_s are the weight factors for RF and SWD datasets (see Real data section for details), and p is a parameter in the range of (0,1) to determine some additional weight.

Then a numerical solver with Leapfrog scheme is applied to solve this first-order partial differential equation systems in equation (4), which could keep the energy conservation property of the Hamiltonian to the first order of the time step (Fichtner et al., 2021). In addition, to tackle the problem-related range of each parameter, we add some constraints in the framework, including hard constraints (like Poisson ratio >0) and soft constraints (like

in some models in which we cannot synthesize our data) in the sampling. The pseudo code of generating a new sample from the current one could be seen in the algorithm chart in Appendix.

One of the problems in HMC sampling is how to properly choose the best hyperparameters to make sure the trajectories of the particle sample the posterior distribution thoroughly. These hyperparameters include the time step length (Δt), the number of steps (L) and even the mass matrix. For Δt and L , they actually determine the distance from the new sample to the current one. If the distance is too small, these two samples cannot be viewed as independent samples. In addition, if the trajectory is too long, it would be not only time consuming, but the trajectory may roll back to the previously sampled region (Neal, 2012). Some techniques are applied to solve this problem such as the NO-U-TURN sampler (Hoffman and Gelman, 2014), but we found that in our problem it will decrease the efficiency significantly. In addition, some other researchers prefer to fix the number of steps (Fichtner et al., 2019). In this article, to avoid periodical trajectories, we randomly choose L and Δt from a given set for every trajectory. This approach was suggested by Neal (2012) and was applied by Aleardi et al. (2020). What is more for the time step Δt is that it not only plays the role like L but also determines the stability of the Leapfrog scheme. It could be shown that the size of it should be less than a threshold (Fichtner et al., 2021). Therefore, in practice, we first test several different Δt s for enough steps to find the approximately upper limit point from stable to unstable. Then we also set soft constraints for this parameter to exclude some outliers in the parameter set, that is, we return to the previous point and change another Δt instead if the searching becomes unstable. The mass matrix controls the truing, regularization, and the particle traveling speed during the sampling (Fichtner et al., 2019), and we will discuss the effect of mass matrix in The choice of mass matrix section.

Forward Method

In the joint inversion framework, we assume that the underground media is composed of a stack of homogenous elastic isotropic layers overlaid on a half-space, that is, $\mathbf{m} = [V_p, V_s, \rho, h]$ for each layer. Because what we need are just the eigenvalues (SWD curves) and the ratio of vertical and horizontal components (RF data) of the wavefield in the frequency domain, and there is no need to consider more general approach like wave-number integration (Zhu and Rivera, 2002). As a result, the matrix propagation method (Haskell, 1953; Dunkin, 1965) is utilized to synthetic RF and SWD data only with different radiation conditions in the half-space. For RF data, we deconvolve it by water level regularization with a water level factor of 0.001.

Derivatives of RF and SWD

One of the requirements for HMC is that we need to evaluate the gradient of the potential energy efficiently, otherwise it will severally affect the efficiency when the number of unknown

increases. Fortunately, the Fréchet kernel of RF and SWD data could be easily computed. Here, we briefly discuss the procedures for calculating these kernels.

For RF data, the surface displacements could be written in the matrix form (Hu and Zhu, 2017):

$$\begin{bmatrix} iA_x \\ A_z \end{bmatrix} = \frac{1}{R_{11}R_{22} - R_{12}R_{21}} \begin{bmatrix} R_{22} & -R_{12} \\ -R_{21} & R_{11} \end{bmatrix} \begin{bmatrix} w_P^U \exp(-v_\alpha z_N) \\ w_{SV}^U \exp(-v_\beta z_N) \end{bmatrix}, \quad (8)$$

in which:

$$R = E_{N+1}^{-1} \mathbf{a}_N \mathbf{a}_{N-1} \dots \mathbf{a}_1. \quad (9)$$

And \mathbf{a}_i is the Thompson–Haskell matrix in layer i , and E_i is the eigenvector matrix of \mathbf{a}_i :

$$E^{-1} = \frac{\gamma}{2} \begin{bmatrix} -1 & -\gamma_1 \frac{k}{v_\alpha} & \frac{1}{2\mu} & \frac{k}{2\mu v_\alpha} \\ \gamma_1 \frac{k}{v_\beta} & 1 & -\frac{k}{2\mu v_\beta} & -\frac{1}{2\mu} \\ 1 & -\gamma_1 \frac{k}{v_\alpha} & -\frac{1}{2\mu} & \frac{k}{2\mu v_\alpha} \\ -\gamma_1 \frac{k}{v_\beta} & 1 & \frac{k}{2\mu v_\beta} & -\frac{1}{2\mu} \end{bmatrix}, \quad (10)$$

in which $\gamma = 2k^2\beta^2/\omega^2$ and $\gamma_1 = 1 - 1/\gamma$. Both E_{N+1}^{-1} and \mathbf{a}_i are composed of elementary function operations of model parameters. $W_{P,SV}^U$ is the upgoing wave component for P/SV wave in the half-space. The RF is defined as

$$F^P = \frac{A_x}{A_z} = \frac{iR_{22}}{R_{21}}, F^{SV} = \frac{A_z}{A_x} = -\frac{iR_{11}}{R_{12}}. \quad (11)$$

Then the derivative of RF for parameter m could be written as

$$F_{,m}^P = i \frac{R_{22,m}R_{21} - R_{21,m}R_{22}}{R_{21}^2}, F_{,m}^{SV} = i \frac{R_{11,m}R_{12} - R_{12,m}R_{11}}{R_{12}^2}. \quad (12)$$

And the derivatives of R could be transferred to Thompson–Haskell matrix

$$\mathbf{R}_{,m} = \begin{cases} E_{N+1,m}^{-1} \mathbf{a}_N \mathbf{a}_{N-1} \dots \mathbf{a}_1, & \text{if } m \text{ in the half-space} \\ E_{N+1}^{-1} \mathbf{a}_N \mathbf{a}_{N-1} \dots \mathbf{a}_{i+1} \mathbf{a}_{i,m} \mathbf{a}_{i-1} \dots \mathbf{a}_1, & \text{if } m \text{ in the } n\text{th space} \end{cases} \quad (13)$$

The concrete form of these expressions could be seen in Hu and Zhu (2017).

For SWD data the Fréchet kernel is a little more cumbersome than RF to find. Here we take the variational approach (Gomberg and Masters, 1988; Aki and Richards, 2002). In the frequency–wavenumber domain, the wave equation in each homogeneous isotropic layer could be expressed as (Aki and Richards, 2002):

$$\begin{cases} -\rho\omega^2 u_i = \lambda \nabla_i \left(\frac{\partial}{\partial x_j} u_j \right) + \mu \nabla^2 u_i \\ T_i|_{z=0} = 0; u_i|_{z \rightarrow \infty} = 0; \end{cases} \quad (14)$$

in which $\nabla_i \rightarrow (ik, 0, \frac{\partial}{\partial z})$, and $c = \omega/k$ is the phase velocity. Then we could write the variational form of this elastodynamics problem:

$$I = \int_0^\infty L(\mathbf{u}, c, \mathbf{m}) dz = 0, \quad (15)$$

in which I is the action, L is the Lagrangian of this problem (Dahlen and Tromp, 1998; Aki and Richards, 2002), and $\delta I = 0$ for all eigenfunction u_i . For example, the Lagrangian of Rayleigh wave is

$$L = \frac{1}{4} \rho \omega^2 (r_1^2 + r_2^2) - \frac{1}{4} \left[\lambda \left(kr_1 + \frac{dr_2}{dz} \right)^2 + \mu \left(\frac{dr_1}{dz} - kr_2 \right)^2 + 2\mu k^2 r_1^2 + 2\mu \left(\frac{dr_2}{dz} \right)^2 \right], \quad (16)$$

in which r_1, r_2 are the displacement eigenfunctions in horizontal and vertical directions, respectively.

Then by perturbation theory we could relate the perturbation of phase velocity to that of the model parameters by

$$0 = \sum_{m=1}^\infty \int_{z_m}^{z_{m+1}} \left(\delta \rho_m \frac{\partial L}{\partial \rho_m} + \delta \alpha_m \frac{\partial L}{\partial \alpha_m} + \delta \beta_m \frac{\partial L}{\partial \beta_m} \right) dz - \sum_{m=1}^\infty [L_m]^+ \delta z_m + \delta c \int_0^\infty \frac{\partial L}{\partial c} dz, \quad (17)$$

in which $[L_i]^\pm$ is the jump of the action across the i th discontinuity. From equation (14), we could obtain the derivatives for any model parameters for a fixed ω by the ratio of two integrals. For example, for the S-wave velocity, we could have

$$\left(\frac{\partial c}{\partial \beta} \right)_m = \left(\frac{\beta_m \rho_m}{UI_0} \right) \int_{z_m}^{z_{m+1}} \left[\left(r_1 + \frac{1}{k} \frac{dr_2}{dz} \right)^2 + \frac{4}{k} r_2 \frac{dr_1}{dz} \right] dz, \quad (18)$$

in which U is the group velocity and

$$I_0 = \int_0^\infty \rho (r_1^2 + r_2^2) dz. \quad (19)$$

A subtle problem is that in equation (14) the Fréchet kernel for the discontinuities should be converted to that of the thickness in our model parameterization. It could be converted by the following formula:

$$K_h^i = \sum_{j=i+1}^N K_{\text{discont}}^j \quad (20)$$

For group velocity, the derivatives could be obtained by utilizing the relation to the phase velocity:

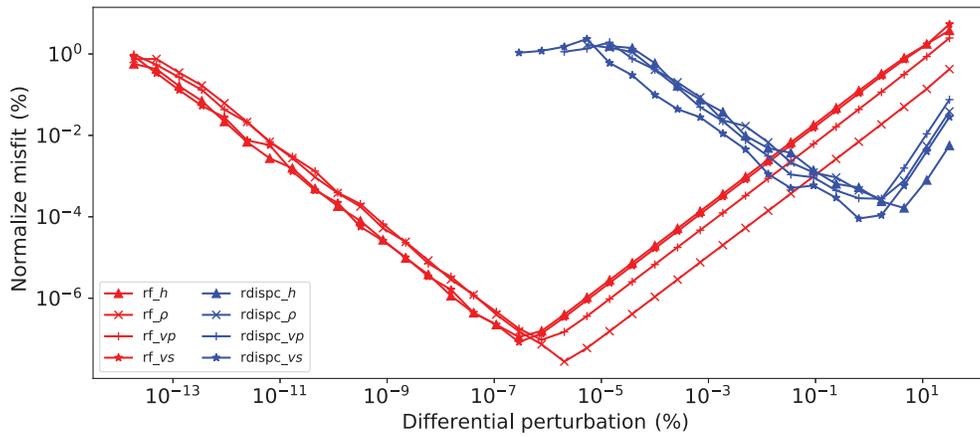


Figure 1. Misfit of analytic and differential method. The misfit of receiver function is in red, and Rayleigh phase velocity is in blue. The Gaussian factor and ray parameter of receiver function (RF) are 1.5 and 0.045 s/km, respectively.

$$\frac{\partial U}{\partial m} = \frac{\partial U}{\partial c} \frac{\partial c}{\partial m} = \frac{U}{c} \left(2 - \frac{U}{c} \right) \frac{\partial c}{\partial m} - \frac{U^2}{c^2} T \frac{\partial}{\partial T} \frac{\partial c}{\partial m}, \quad (21)$$

in which T is the period, and $\frac{\partial}{\partial T} \frac{\partial c}{\partial m}$ term could be evaluated by central-difference approach.

Here, we list an example to show the accuracy of the Fréchet kernels. We calculate the Fréchet kernel of P -RF and Rayleigh phase velocity by analytic and differential method with the 7-layer model shown in synthetic test. The range of differential perturbation from $10^{-9}\%$ to 10% for α (P -wave velocity), β (S -wave velocity), ρ (density), and h (thick of layer). Figure 1 shows that the discrepancy between analytic and differential derivatives is under 1% for common difference step.

Application Synthetic test

We apply our algorithm to synthetic data set to verify its validity and limitation. Here, we conduct four types of tests. First, we set up four models to perform the inversion. The first model is a simple 3-layer model with the Moho interface at 40 km and

a layer in the lithosphere at 70 km. The velocity is given according to the AK135 model (Kennett *et al.*, 1995). The second model is the 4-layer model with a 5 km thick sedimentary layer at the top with 1.9 km/s shear wavespeed. In the third model, the crust is divided into the upper and lower crust, and there is a 5 km thick low-velocity layer (3.3 km/s) in the lower crust. The last model is a double low-velocity model that contains a low-velocity layer in the upper and lower crust. All the inversions are performed with 400 trials (including 100 trials in the burn-in

phase) for every 16 chains. The searching range is $\pm 20\%$ around the true model, including the velocity and thickness of each layer.

To demonstrate the advantage of our algorithm, we compare our algorithm with rj-McMC method (Dreiling and Tilmann, 2019) with 16 chains (same as HMC method). We choose different numbers of samplings to make sure similar computation time for each experiment; the statistics are shown in Table 1. Figure 2 shows that the Moho interface is recovered well in all the models, and same as the structure such as shallow sedimentary and low-velocity layers, especially the low-velocity of 30 km in Figure 2c. The acceptance rate of the HMC method is about 90%. Meanwhile, the acceptance rate of the rj-McMC is only around 35%~40%. The misfit curve in Figure 3 indicates that the HMC method will converge to a stationary regime much faster in only 20 samples in the synthetic test.

Second, we choose a different searching range, from $\pm 10\%$ to $\pm 65\%$ around the true model with the 7-layer model in 400 trials to test how the accuracy of the prior model will affect the inversion. The result of Figure 4 shows that the more accurate

TABLE 1
Parameter and Computation Time of Different Model Inversion

Model Name	HMC (This Study)			rj-McMC (Bayhunter)		
	Time (s)	Parameter	Samples	Time (s)	Parameter	Samples
Model 1	33.04	$3 \times 2 = 6$	400	34.69	2–4	11,000
Model 2	77.77	$4 \times 2 = 8$		78.31	3–5	20,000
Model 3	100.60	$5 \times 2 = 10$		103.27	4–6	25,000
Model 4	129.59	$7 \times 2 = 14$		132.24	6–8	25,000

Note that the velocity and thickness are inverted simultaneously in HMC method so that the actual parameter is double. HMC, Hamiltonian Monte Carlo; and rj-McMC, reversible-jump Markov chain Monte Carlo.

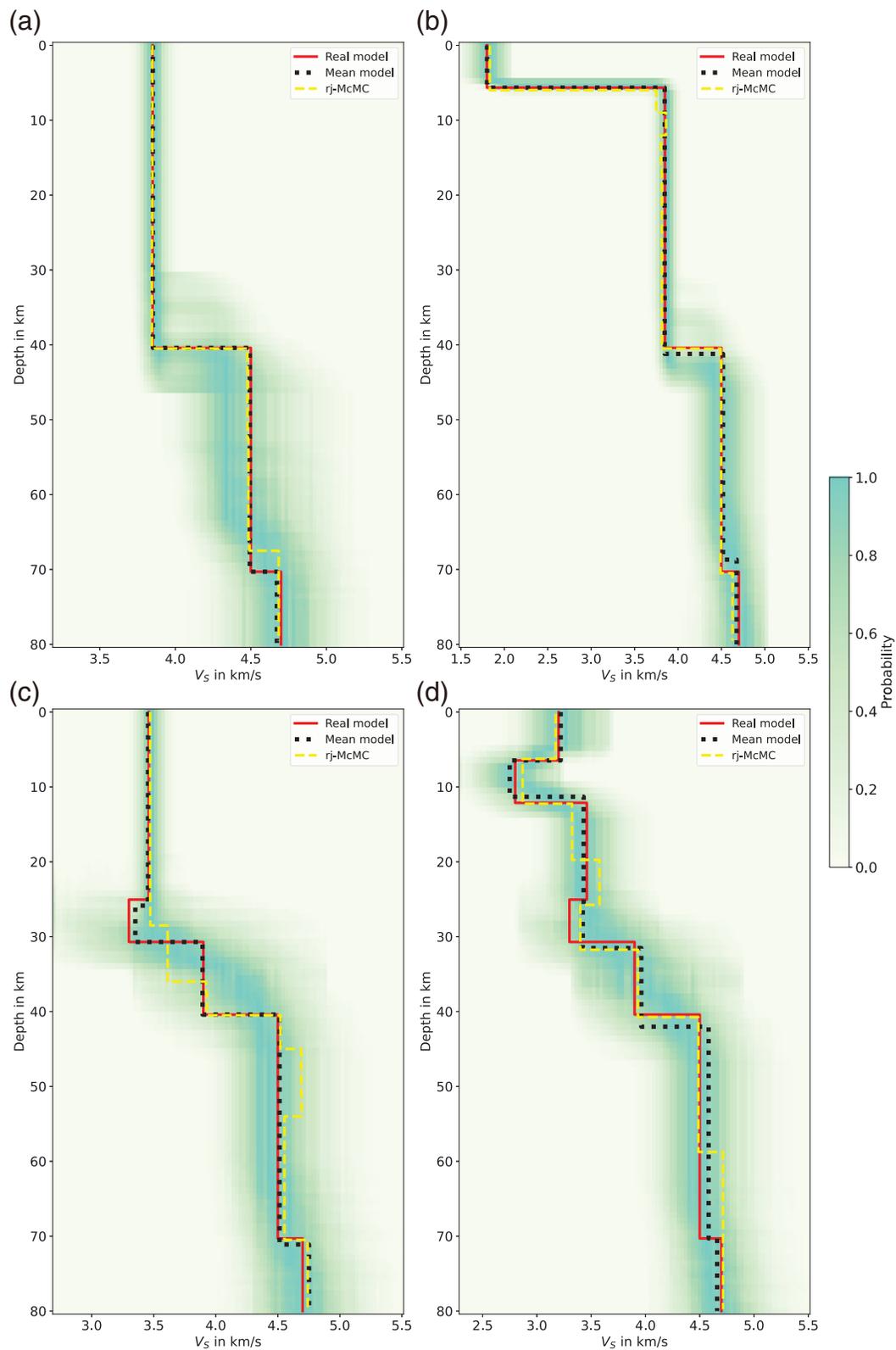


Figure 2. Inversion results of different models with Hamiltonian Monte Carlo (HMC) framework. (a) Simple 3-layer model; (b) 4-layer model with sedimentary; (c) 5-layer model with single low-velocity layer; and (d) 7-layer model with double low-velocity

layer. The red line is the true model, and black dotted line is the mean model, which is average of 1% best (minimum misfit) model. The blue dashed line is the mean model, which is inverted by reversible-jump Markov chain Monte Carlo (rj-McMC; Bayhunter).

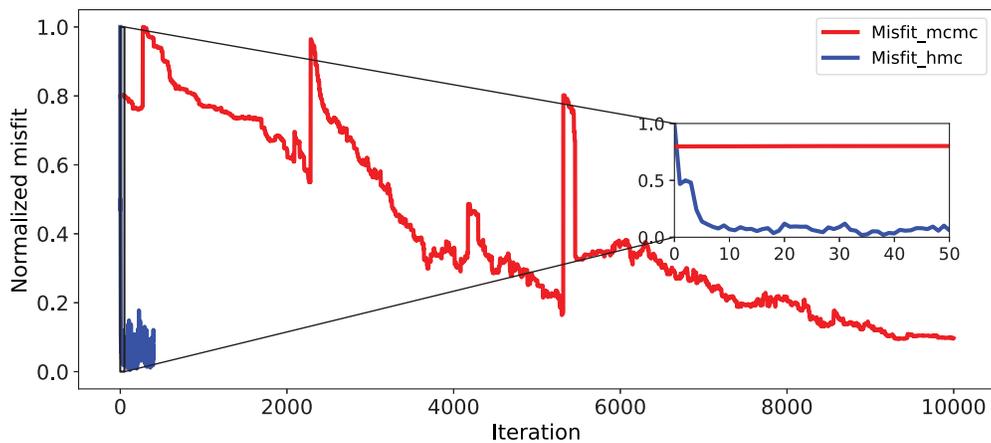


Figure 3. The misfit curve of model 4 inversion. The red line is the Markov chain Monte Carlo (MCMC) method, and only the first 10,000 iterations are present. The blue line is the HMC method. The closeup of the first 50 iterations is present in the figure.

prior model will provide a better result. In most cases, the searching range should be around $\pm 20\% \sim \pm 40\%$ of the true model for our algorithm. Third, we test different numbers of sampling for inverting the 7-layer model within a $\pm 40\%$ searching range. Figure 5 indicates that the more trials we perform the better result we get. The computation cost will tremendously increase for calculating gradient, and 400 trials are enough for this synthetic test.

Finally, we try different initial model to check whether our algorithm relies on it. Different from the previous synthetic test to which initial model is drawn randomly, we deliberately cast the initial model evenly in the range of $\pm 30\%$ with 800 trials. Figure 6 shows that the Moho discontinuity is recovered well, while there is still deviation of second layer (70 km depth interface). The velocity is relatively accurate (~ 4.7 km/h). The deviation may be caused by the low-velocity gradient of second layer (from 4.5 to 4.7 km/h) and the wide search range of depth (from 49 to 91 km). Therefore, we only need to set evenly distributed initial models and then remove those outlier chains after inversion.

Real data

We also apply our algorithm to a real dataset—a borehole station IU.CHTO, which was located in Thailand for more than 10 yr of operation. We perform the inversion with 800 trials and set the weight factors $w_{rf} = w_{swd} = 1$. The weight factors are chosen according by the data noise of data set. For general cases, the data noise could be estimated by statistical method such as Bootstrapping (Zoubir and Iskander, 2004). The model setting could be found in Table A1. The result in Figure 7 indicates the smooth Moho interface from the depth at 25 to 35 km, and shear wavespeed increases from 3.4 to 4.5 km/s, which agreed with the previous studies (Wang et al., 2018). Majority $H - \kappa$ result (Bai et al., 2010; Noisagool et al., 2014) shows that the Moho layer of IU.CHTO is around

30 km, which is close to the center of our Moho result. There is a negative phase around 8.5 s, which we did not fit well. According to the stacked RF profile (Bai et al., 2010; Noisagool et al., 2014), this negative phase only occurs in RF with high ray parameter (above 0.0625 s/km) and northeast direction. It may indicate the anisotropy of underneath medium. The inversion result with higher Gauss factor of RF also leads to the same interpretation (see the Fig. A1). Compared with the velocity model of other research, the discrepancy

of Moho is mainly due to the limitation of layer number, and the deviation of shallow layer might cause by the lacking of HV ratio data. Although only one RF and Rayleigh wave dispersion are used, the principal feature is recovered in few trials.

Discussion

The fixed-dimension dilemma

Since the HMC is a kind of fix-dimensional method, we have to choose the hyperparameter (the number of layers) before performing the joint inversion. Although the number of layers is a hyperparameter in this HMC method, there are still several ways to choose the number of layers. First, Carter et al. (2021) suggest that we could perform other methods such as rj-McMC or normal linear inversion methods and then choose the parameter based on the result. As long as we care not to introduce the possible bias, we could achieve an appropriate estimation of the number of layers. Second, Aleari et al. (2020) point out that the dimensionality could be decided by χ^2 or Bayesian information criterion (BIC) (Schwarz, 1978; Sambridge et al., 2006; Ando, 2010). The BIC is defined as

$$\text{BIC} = k \ln n - 2 \ln L, \quad (22)$$

in which L denotes the maximized value of the likelihood function of the model, n is the number of data points, and k is the number of model parameters. Figure 8 illustrates the result what if we use a wrong number of layers to perform the HMC inversion for model 4, and Table 2 presents the statics result of synthetic.

Combining Table 2 and Figure 8, we could conclude that although the misfit for larger hyperparameters is lower, it will also obtain more penalty. The BIC of the 7-layer is lower than other models, so we can choose the appropriate hyperparameter.

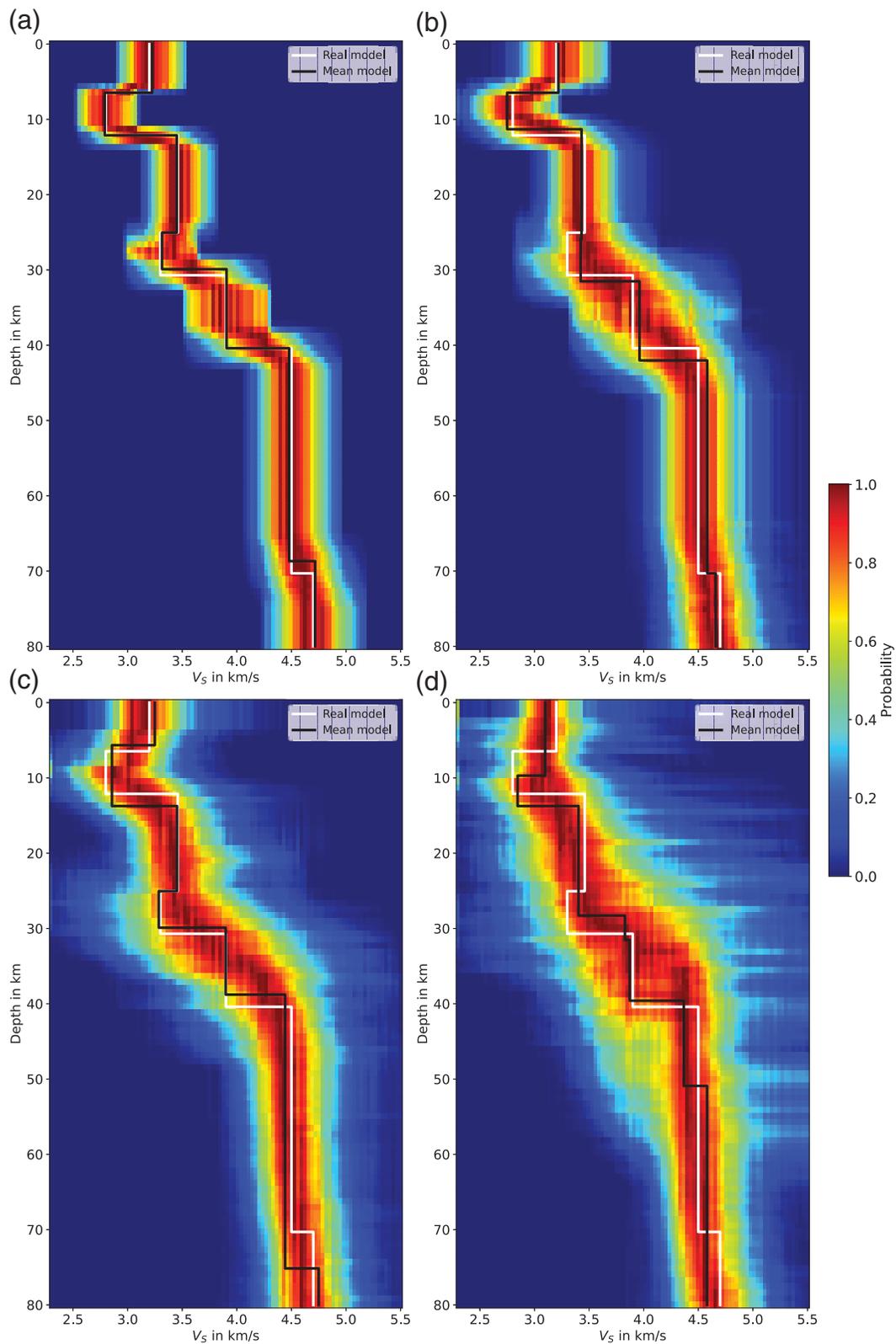
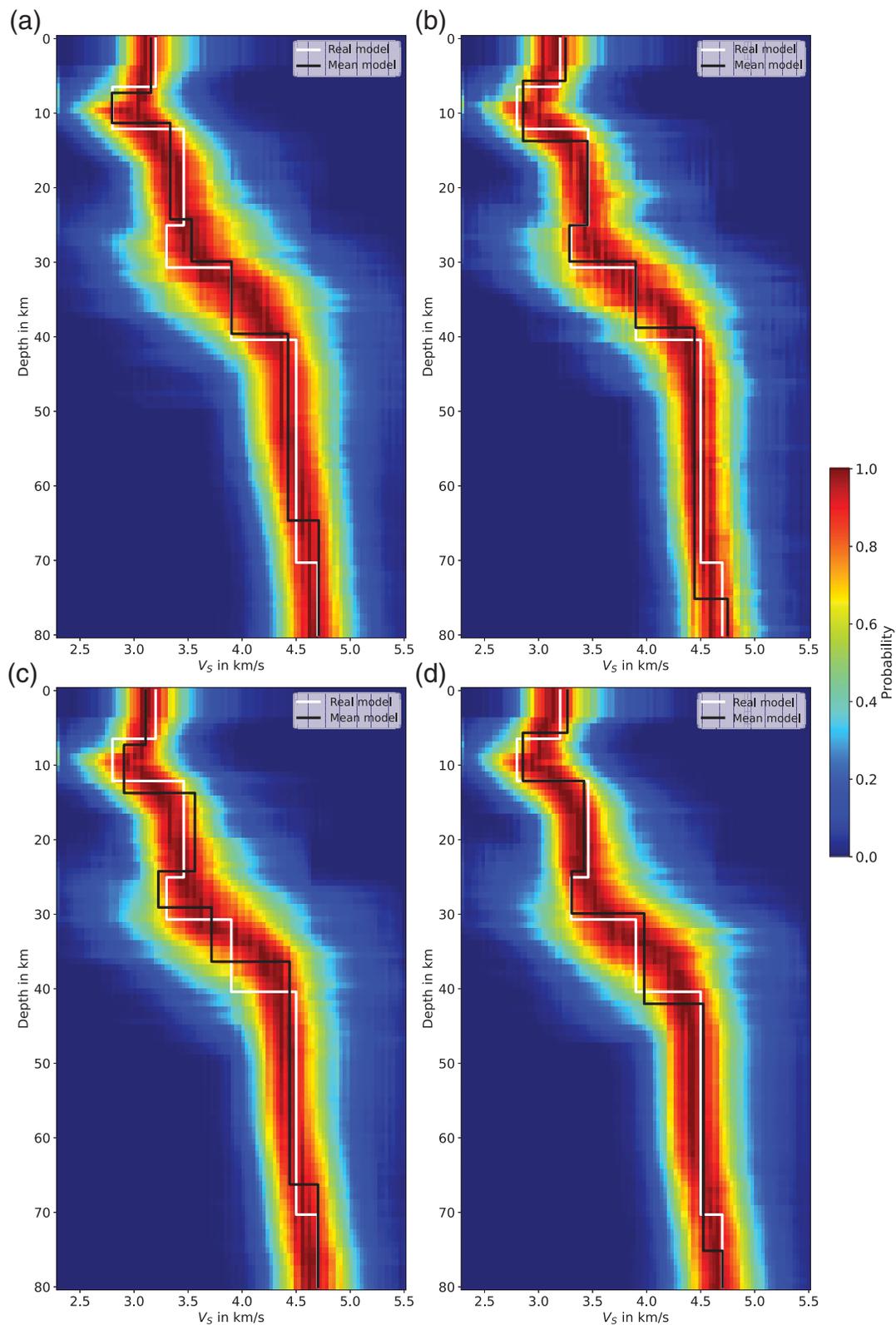


Figure 4. Inversion results of different searching range with 7-layer model. (a) $\pm 10\%$ searching range; (b) $\pm 20\%$ searching range; (c) $\pm 40\%$ searching range; and (d) $\pm 65\%$ searching range.

The choice of mass matrix

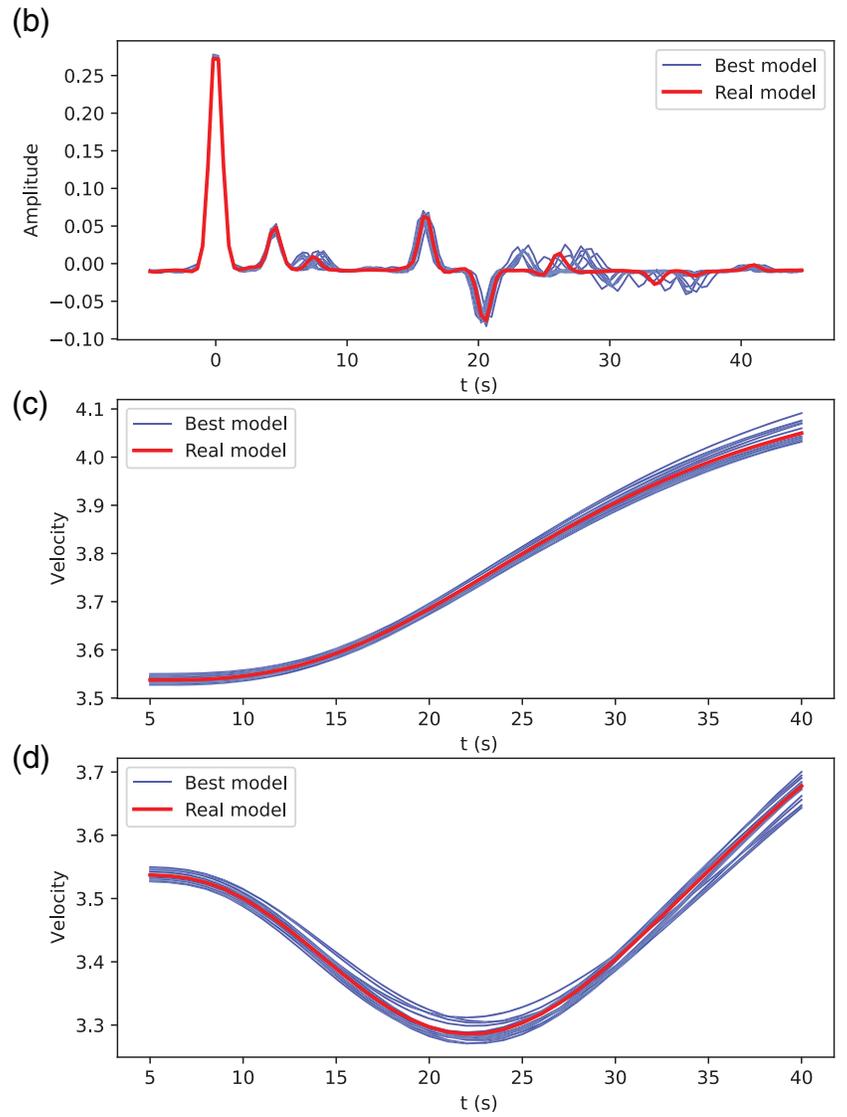
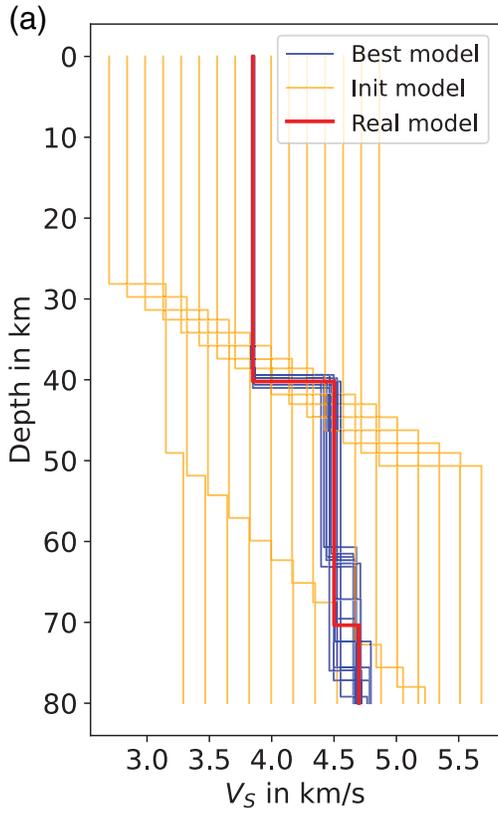
The mass matrix plays an important role during the HMC sampling. The common mass matrix is proportional to the identical matrix I . In our research, we found that the gradient of velocity is larger than the thickness. So, the particle will



oscillate in the model space of velocity rapidly, and the sampling processing in the model space of thickness may be insufficient.

To tune the Hamiltonian trajectory of particle, we could substitute the mass matrix by

Figure 5. Inversion results of different trials. (a) 200 trials; (b) 400 trials; (c) 800 trials; and (d) 1000 trials.



$$\mathbf{M} = \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \mathbf{C}_m^{-1}, \quad (23)$$

in which \mathbf{G} is the linear forward operator, \mathbf{C}_d is the data covariance, and \mathbf{C}_m is the prior model covariance (Fichtner *et al.*, 2019). However, this mass matrix is only appropriate for linear problems, and the inverse of mass matrix equation (23) may not have explicit forms. Therefore, to capture the local gradient information, we introduce the Hessian matrix \mathbf{H} of potential energy in Hamiltonian system (Fu *et al.*, 2016). The problem becomes the estimation of Hessian matrix. The L-BFGS method is the most popular quasi-Newton method. In L-BFGS approximation (Nocedal and Wright, 1999), the inverse Hessian matrix for \mathbf{m}_k will be

$$\mathbf{H}_{k+1} = \left(\mathbf{I} - \frac{\Delta \mathbf{g}_k \Delta \mathbf{m}_k^T}{\Delta \mathbf{m}_k^T \Delta \mathbf{g}_k} \right) \mathbf{H}_k \left(\mathbf{I} - \frac{\Delta \mathbf{g}_k \Delta \mathbf{m}_k^T}{\Delta \mathbf{m}_k^T \Delta \mathbf{g}_k} \right) + \mathbf{m}_k \mathbf{m}_k^T, \quad (24)$$

in which $\Delta \mathbf{m}_k = \mathbf{m}_{k+1} - \mathbf{m}_k$ and $\Delta \mathbf{g}_k = \nabla U(\mathbf{m}_{k+1}) - \nabla U(\mathbf{m}_k)$. Figure 9 shows that the particle moves much longer

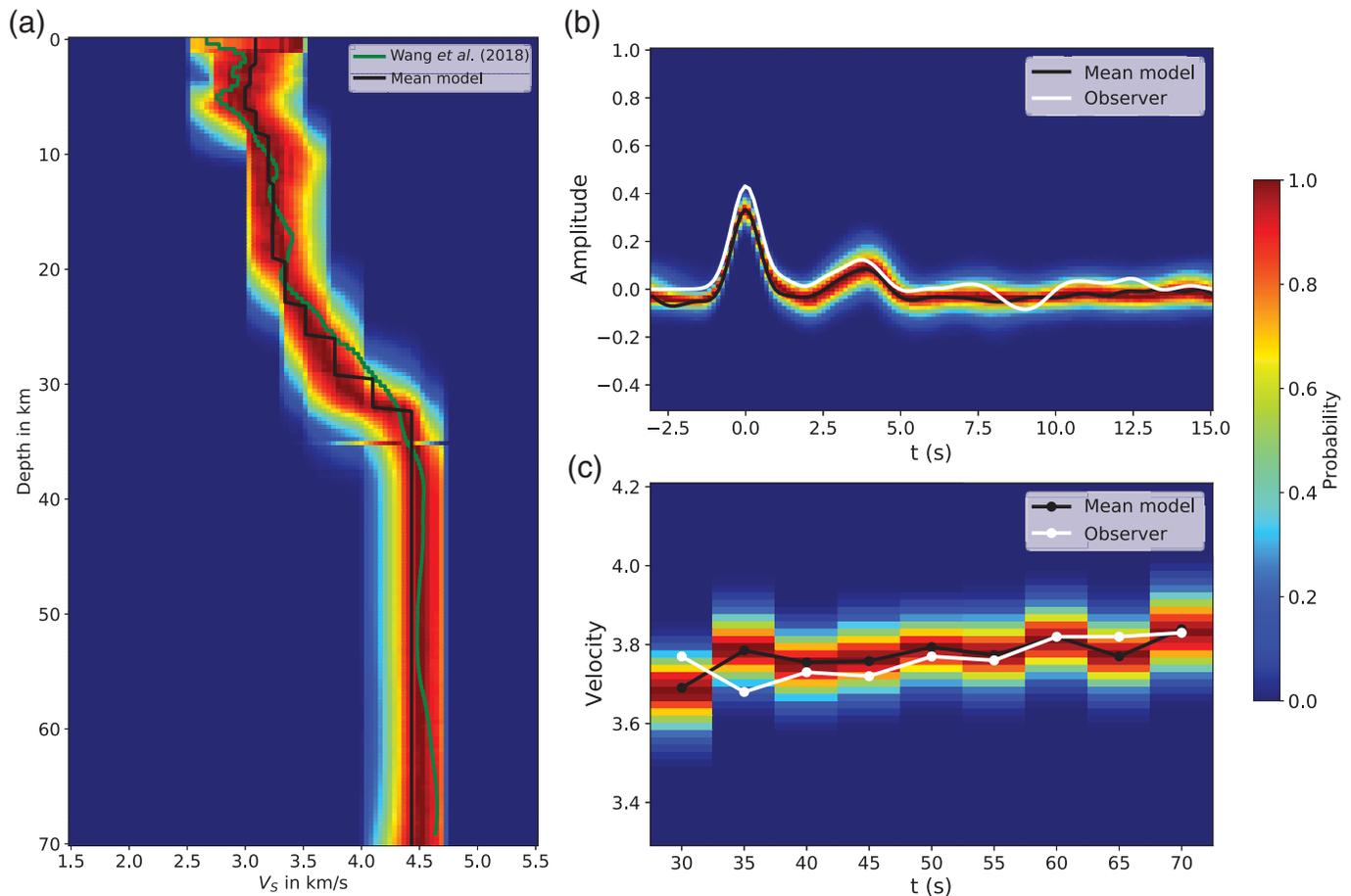
Figure 6. Inversion results of different initial models. (a) Inversion result; (b) receiver function fitting; (c) phase velocity of Rayleigh wave fitting; and (d) group velocity of Rayleigh wave fitting. The gray lines are the best (the minimum misfit) results of each chain, and yellow lines are different initial model.

in the model space of thickness after substituting the mass matrix.

In addition, there are still many choices for mass matrix, such as decrease with the iteration (Lima *et al.*, 2022). There may not exist a universal choice, and these choices may depend on the specific properties of the inversion problems.

Is the prior strong?

Compared to other rj-McMC methods, such as the research from Dreiling and Tilmann (2019) or Bai *et al.* (2021), we acknowledge that the prior for first synthetic test (Fig. 2)



seems relatively strong. Although the rj-McMC method allows a larger range of prior, this advantage is usually based on massive chains, trails, and strictly selection criterion. For example, Dreiling and Tilmann (2019), also the authors of Bayhunter, use a wider range (V_S from 2~5 km/h, V_p/V_S from 1.45~2.05, and the maximum number of layers is 20) priors for rj-McMC method. The inversion was performed with 100 chains, and each chain performed 1.8 million iterations, and those chains for which likelihood function is significantly lower than other chains were declared as outliers and will not be considered in choosing the inverted model. Bai et al. (2021) explored the model space 2 million times with the 3:2 ratio for burn-in and exploration phase. To eliminate the unreasonable solution, the outlier chains for which likelihood functions are lower than 0.8 times the median of all chains were discarded. Moreover, the model that has a shear velocity of greater than 4.0 km/s or less than 3.0 km/s in bottom half-space is also rejected. Those processing could be considered as a kind of prior in practical.

We perform the inversion of model 4 with 50 chains and 1200 trials (see Fig. A2). The prior of v_s is uniform distribution in range 2.5~5 km/s, and the prior of thickness is $\pm 50\%$ around the true value, which might be considered as a noninformative prior. After we discard the outlier, the model is also recovered well except the fifth layer.

Figure 7. Inversion result of IU.CHTO with HMC framework. (a) Inversion result; (b) receiver function fitting; and (c) phase velocity of Rayleigh wave fitting.

TABLE 2

The Bayesian Information Criterion (BIC) and χ^2 for Different Parameters in Two Synthetic Tests

Synthetic Test	Parameters	BIC	χ^2
Model 4 (7 layer)	6 layers	81.171	0.1255
	7 layers	74.344	0.1082
	8 layers	81.732	0.1002

For many linearized joint inversions, the difference between initial model and the final result is just around 20%, such as the research of Tang et al. (2022) or Döring et al. (2022). We believe that it is accurate enough to set the prior based on local 1D reference model and CRUST 1.0.

Conclusion

We proposed a Hamiltonian Monte Carlo sampling method for the joint inversion of RF and SWD in this article. The

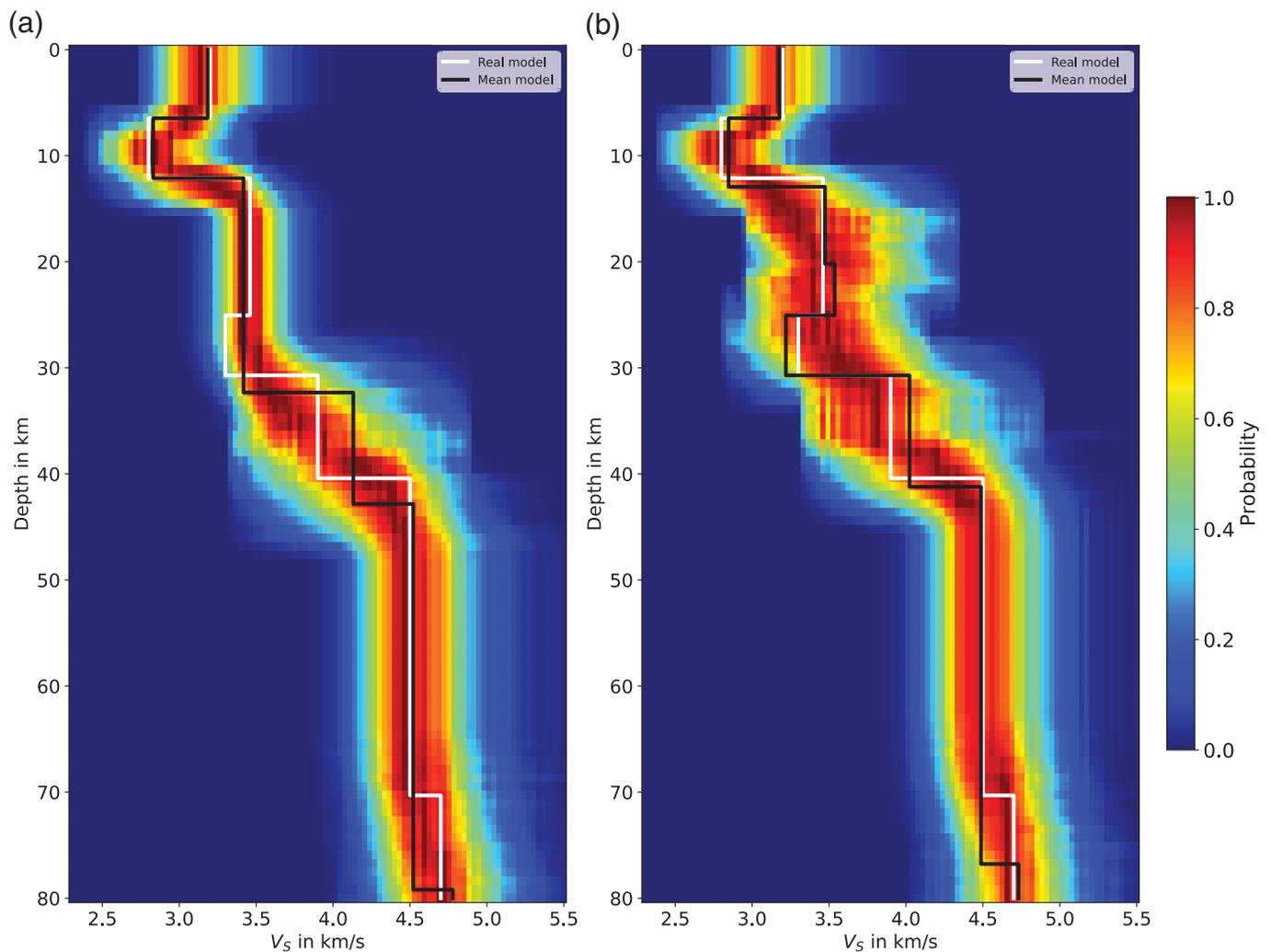


Figure 8. The inversion result of model 4 with different number of layers. (a) 6-layers result and (b) 8-layers result.

synthetic and real data test shows the applicability and usefulness of our HMC algorithm. The number of required samples reduces dramatically by introducing the gradient information during the inversion. We applied a semianalytical method to efficiently compute the gradient of the misfit function of this inverse problem, which gives an advantage of achieving the tradeoff between a high acceptance rate and a low number of samples. Although the detail of our algorithm can be improved, such as by adding the reverse jump process to determine the number of layers (Sen and Biswas, 2017) or by treating the noise levels as unknown parameters, the results are still promising. We could expect that with the combination of other methods our approach can hope to decode the mysteries of our earth system economically and effectively in the future.

Data and Resources

The data used in this article are the RF and Rayleigh-wave dispersion of IU.CHTO, which is published in Wang *et al.* (2018). The continuous waveform can be obtained from the Incorporated Research Institutions for Seismology (IRIS) Data Management Center. The code used in this article is available at <https://github.com/nqdu/RfSurfHmc> (last accessed June 2022).

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

The authors are appreciated to Longlong Wang and Yuanfang Li for constructive discussions. The data and the inversion result of IU.CHTO are provided by Xin Wang. The inspiring thoughtful discussion with Tianyu Zheng, Mingming Jiang, and Xin Wang had profound influence during the development of this article. The authors thank the constructive comments from the editor and two anonymous reviewers. The authors also appreciate Yuanxin Hao for providing the technical support for teleworking during the strictly self-quarantine in the Covid-19 epidemic. This work is supported by the National Natural Science Foundation of China (Grant Numbers 91958209 and 41974057).

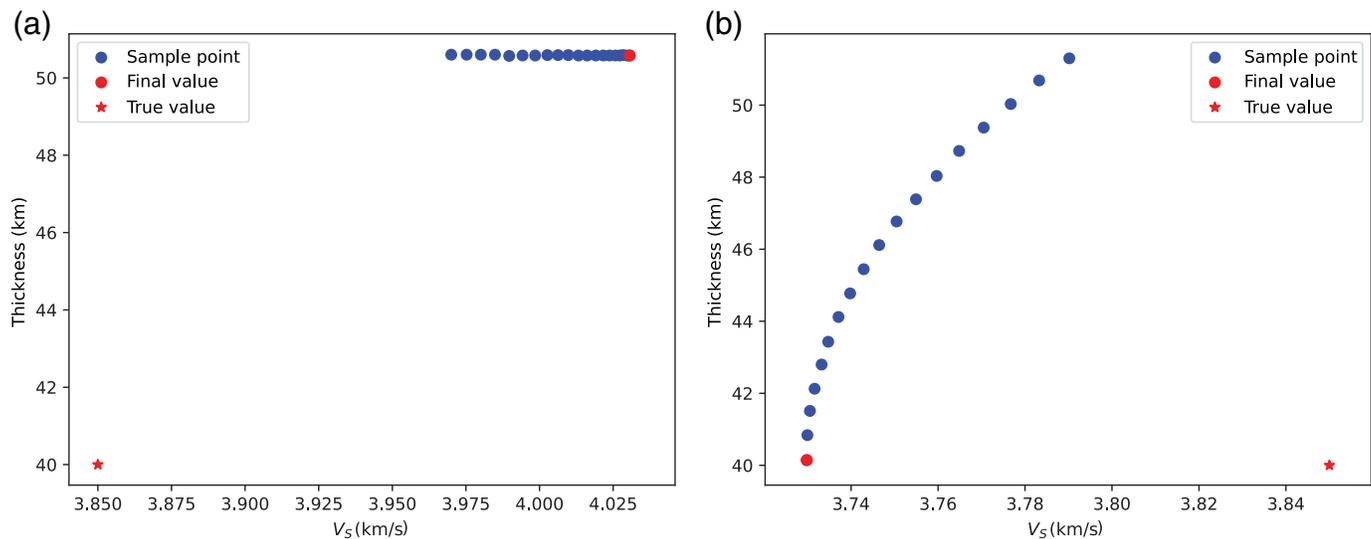


Figure 9. Hamiltonian trajectory of different initial test. (a) The trajectory with identical mass matrix and (b) the trajectory after

substituting the mass matrix with Hessian matrix by L-BFGS method.

References

- Agostinetti, N. P., and A. Malinverno (2010). Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.* **181**, no. 2, 858–872.
- Aki, K., and P. G. Richards (2002). *Quantitative Seismology*, University ScienceBooks, California.
- Aleardi, M., A. Salusti, and S. Pierini (2020). Transdimensional and Hamiltonian Monte Carlo inversions of Rayleigh-wave dispersion curves: A comparison on synthetic datasets, *Near Surf. Geophys.* **18**, no. 5, 515–543.
- Ammon, C. J., G. E. Randall, and G. Zandt (1990). On the nonuniqueness of receiver function inversions, *J. Geophys. Res.* **95**, no. B10, 15,303–15,318.
- Ando, T. (2010). *Bayesian Model Selection and Statistical Modeling*, CRC Press, New York.
- Aster, R. C., B. Borchers, and C. H. Thurber (2012). *Parameter Estimation and Inverse Problems*, Elsevier, New York.
- Backus, G. E., and J. F. Gilbert (1967). Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. Roy. Astron. Soc.* **13**, nos. 1/3, 247–276.
- Bai, L., X. Tian, and J. Ritsema (2010). Crustal structure beneath the Indochina peninsula from teleseismic receiver functions, *Geophys. Res. Lett.* **37**, no. 24, doi: [10.1029/2010GL044874](https://doi.org/10.1029/2010GL044874).
- Bai, Y., Y. He, X. Yuan, F. Tilmann, Y. Ai, M. Jiang, G. Hou, C. T. Mon, M. Thant, and K. Sein (2021). Seismic structure across central Myanmar from joint inversion of receiver functions and Rayleigh wave dispersion, *Tectonophysics* **818**, 229,068.
- Bodin, T., and M. Sambridge (2009). Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.* **178**, no. 3, 1411–1436.
- Bodin, T., M. Sambridge, N. Rawlinson, and P. Arroucau (2012). Transdimensional tomography with unknown data noise, *Geophys. J. Int.* **189**, no. 3, 1536–1556.
- Bodin, T., M. Sambridge, H. Tkalcic, P. Arroucau, K. Gallagher, and N. Rawlinson (2012). Transdimensional inversion of receiver functions and surface wave dispersion, *J. Geophys. Res.* **117**, no. B2, doi: [10.1029/2011JB008560](https://doi.org/10.1029/2011JB008560).
- Bourjot, L., and B. Romanowicz (1992). Crust and upper mantle tomography in Tibet using surface-waves, *Geophys. Res. Lett.* **19**, no. 9, 881–884.
- Carter, S., X. Xie, L. Zhou, and L. Yan (2021). Hybrid Monte Carlo 1-D joint inversion of LOTEM and MT, *J. Appl. Geophys.* **194**, 104,424.
- Chib, S., and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm, *Am. Stat.* **49**, no. 4, 327–335.
- Dahlen, F., and J. Tromp (1998). *Theoretical Global Seismology*, Princeton University Press, Princeton, New Jersey.
- Döring, M., J. Julià, and M. Evain (2022). Joint inversion of receiver functions and surface wave dispersion in the Recôncavo–Tucano basin of NE Brazil: Implications for basin formation, *Geophys. J. Int.* **230**, no. 1, 317–333.
- Dreiling, J., and F. Tilmann (2019). BayHunter-McMCMC transdimensional Bayesian inversion of receiver functions and surface wave dispersion, doi: [10.5880/GFZ.2.4.2019.001](https://doi.org/10.5880/GFZ.2.4.2019.001).
- Dunkin, J. W. (1965). Computation of modal solutions in layered, elastic media at high frequencies, *Bull. Seismol. Soc. Am.* **55**, no. 2, 335–358.
- Fichtner, A., and S. Simute (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media, *J. Geophys. Res.* **123**, no. 4, 2984–2999.
- Fichtner, A., and A. Zunino (2019). Hamiltonian nullspace shuttles, *Geophys. Res. Lett.* **46**, no. 2, 644–651.
- Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.* **175**, no. 2, 665–685.

- Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.* **179**, no. 3, 1703–1725.
- Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. Villasenor (2013). Multiscale full waveform inversion, *Geophys. J. Int.* **194**, no. 1, 534–556.
- Fichtner, A., A. Zunino, and L. Gebraad (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.* **216**, no. 2, 1344–1363.
- Fichtner, A., A. Zunino, L. Gebraad, and C. Boehm (2021). Autotuning Hamiltonian Monte Carlo for efficient generalized nullspace exploration, *Geophys. J. Int.* **227**, no. 2, 941–968.
- Fu, T., L. Luo, and Z. Zhang (2016). Quasi-Newton Hamiltonian Monte Carlo, UAI.
- Gebraad, L., C. Boehm, and A. Fichtner (2020). Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. Geophys. Res.* **125**, no. 3, e2019JB018428.
- Girolami, M., and B. Calderhead (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods, *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **73**, no. 2, 123–214.
- Gomberg, J. S., and T. G. Masters (1988). Waveform modeling using locked-mode synthetic and differential seismograms—Application to determination of the structure of Mexico, *Geophys. J. Oxf.* **94**, no. 2, 193–218.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**, no. 4, 711–732.
- Haskell, N. A. (1953). The dispersion of surface waves on multilayered media, *Bull. Seismol. Soc. Am.* **43**, no. 1, 17–34.
- Hastings, W. K. (1970). Monte-Carlo sampling methods using Markov Chains and their applications, *Biometrika* **57**, no. 1, 97–109.
- Hoffman, M. D., and A. Gelman (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, *J. Mach. Learn. Res.* **15**, no. 1, 1593–1623.
- Hu, S. Q., and L. P. Zhu (2017). Calculation of differential seismograms using analytic partial derivatives—I: Teleseismic receiver functions, *Geophys. J. Int.* **210**, no. 2, 887–897.
- Jackson, D. D. (1979). The use of a priori data to resolve non-uniqueness in linear inversion, *Geophys. J. Int.* **57**, no. 1, 137–157.
- Julià, J., C. J. Ammon, R. B. Herrmann, and A. M. Correig (2000). Joint inversion of receiver function and surface wave dispersion observations, *Geophys. J. Int.* **143**, no. 1, 99–112.
- Kennett, B. L., E. Engdahl, and R. Buland (1995). Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.* **122**, no. 1, 108–124.
- Langston, C. A. (1979). Structure under Mount Rainier, Washington, inferred from teleseismic body waves, *J. Geophys. Res.* **84**, no. B9, 4749–4762.
- Lima, P. D. S., G. Corso, M. S. Ferreira, and J. M. de Araújo (2022). Acoustic full waveform inversion with Hamiltonian Monte Carlo method, available at <https://arxiv.org/abs/2206.00622>.
- Muir, J. B., and H. Tkalcic (2020). Probabilistic lowermost mantle P-wave tomography from hierarchical Hamiltonian Monte Carlo and model parametrization cross-validation, *Geophys. J. Int.* **223**, no. 3, 1630–1643.
- Nawaz, M. A., and A. Curtis (2018). Variational Bayesian inversion (VBI) of quasi-localized seismic attributes for the spatial distribution of geological facies, *Geophys. J. Int.* **214**, no. 2, 845–875.
- Neal, R. M. (2012). MCMC using Hamiltonian dynamics, available at <https://arxiv.org/abs/1206.1901>.
- Nocedal, J., and S. J. Wright (1999). *Numerical Optimization*, Springer, New York.
- Noisagool, S., S. Boonchaisuk, P. Pornsopin, and W. Siripunvaraporn (2014). Thailand's crustal properties from tele-seismic receiver function studies, *Tectonophysics* **632**, 64–75.
- Nunn, C., S. W. Roecker, K. F. Priestley, X. F. Liang, and A. Gilligan (2014). Joint inversion of surface waves and teleseismic body waves across the Tibetan collision zone: The fate of subducted Indian lithosphere, *Geophys. J. Int.* **198**, no. 3, 1526–1542.
- Rawlinson, N., and M. Sambridge (2004). Wave front evolution in strongly heterogeneous layered media using the fast marching method, *Geophys. J. Int.* **156**, no. 3, 631–647.
- Rawlinson, N., and M. Sambridge (2005). The fast marching method: An effective tool for tomographic imaging and tracking multiple phases in complex layered media, *Explor. Geophys.* **36**, no. 4, 341–350.
- Sambridge, M. (1999). Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.* **138**, no. 2, 479–494.
- Sambridge, M., and K. Mosegaard (2002). Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.* **40**, no. 3, 3–1–3–29.
- Sambridge, M., K. Gallagher, A. Jackson, and P. Rickwood (2006). Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.* **167**, no. 2, 528–542.
- Schwarz, G. (1978). Estimating the dimension of a model, *Ann. Stat.* **6**, 461–464.
- Sen, M. K., and R. Biswas (2017). Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm, *Geophysics* **82**, no. 3, R119–R134.
- Sen, M. K., R. Biswas, P. Mandal, and P. Kumar (2014). Basis pursuit receiver function, *Bull. Seismol. Soc. Am.* **104**, no. 6, 2673–2682.
- Shen, W. S., M. H. Ritzwoller, V. Schulte-Pelkum, and F. C. Lin (2013). Joint inversion of surface wave dispersion and receiver functions: A Bayesian Monte-Carlo approach, *Geophys. J. Int.* **192**, no. 2, 807–836.
- Siahkoochi, A., G. Rizzuti, and F. J. Herrmann (2020). *Uncertainty Quantification in Imaging and Automatic Horizon Tracking—A Bayesian Deep-Prior Based Approach*, SEG Technical Program Expanded Abstracts 2020, Society of Exploration Geophysicists, 1636–1640.
- Tang, Z., J. Julià, P. M. Mai, W. D. Mooney, and Y. Wu (2022). Shear-wave velocity structure beneath northeast china from joint inversion of receiver functions and Rayleigh wave phase velocities: implications for intraplate volcanism, *J. Geophys. Res.* **127**, no. 5, e2022JB023956 .
- Tarantola, A. (1984). Inversion of seismic-reflection data in the acoustic approximation, *Geophysics* **49**, no. 8, 1259–1266.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia.
- Thurber, C., and J. Um (1987). A fast algorithm for two-point seismic ray tracing, *Bull. Seismol. Soc. Am.* **77**, no. 3, 972–986.

Wang, X., S. Wei, Y. Wang, P. M. Maung, J. Hubbard, P. Banerjee, B. S. Huang, K. Moe Oo, T. Bodin, and A. Foster (2018). A 3D shear-wave velocity model for Myanmar region, *J. Geophys. Res.* **124**, no. 1, 504–526.

West, M., W. Gao, and S. Grand (2004). A simple approach to the joint inversion of seismic body and surface waves applied to the southwest U.S., *Geophys. Res. Lett.* **31**, no. 15, doi: [10.1029/2004GL020373](https://doi.org/10.1029/2004GL020373).

Zhang, X., and A. Curtis (2020). Seismic tomography using variational inference methods, *J. Geophys. Res.* **125**, no. 4, e2019JB018589.

Zhang, X., M. A. Nawaz, X. Zhao, and A. Curtis (2021). An introduction to variational inference in geophysical inverse problems, *Adv. Geophys.* **62**, 73–140.

Zhao, L. F., and S. M. Mousavi (2018). Lateral variation of crustal Lg attenuation in eastern North America, *Sci. Rep.* **8**, no. 1, 7285.

Zhao, L. F., X. B. Xie, J. K. He, X. B. Tian, and Z. X. Yao (2013). Crustal flow pattern beneath the Tibetan plateau constrained by regional Lg-wave Q tomography, *Earth Planet. Sci. Lett.* **383**, 113–122.

Zhu, L. P., and L. A. Rivera (2002). A note on the dynamic and static displacements from a point source in multilayered media, *Geophys. J. Int.* **148**, no. 3, 619–627.

Zoubir, A. M., and D. R. Iskander (2004). *Bootstrap Techniques for Signal Processing*, Cambridge University Press, Cambridge.

Appendix

This appendix provides the details related to the algorithm of one-step HMC sampling, the Table A1 for the model setting of real data inversion, and inversion result of higher Gauss factor data and wider search range, respectively (Figs. A1 and A2).

The Inversion result of higher Gauss factor in real data test

To verify the inversion result of IU.CHTO, we also perform the inversion with higher Gauss factor of receiver function. The data were obtained by EarthScope Automated Receiver Survey. Compared with the result of lower Gauss factor, the low-velocity layer is clearer, and the posterior probability is narrower. The inversion result is in Figure A1.

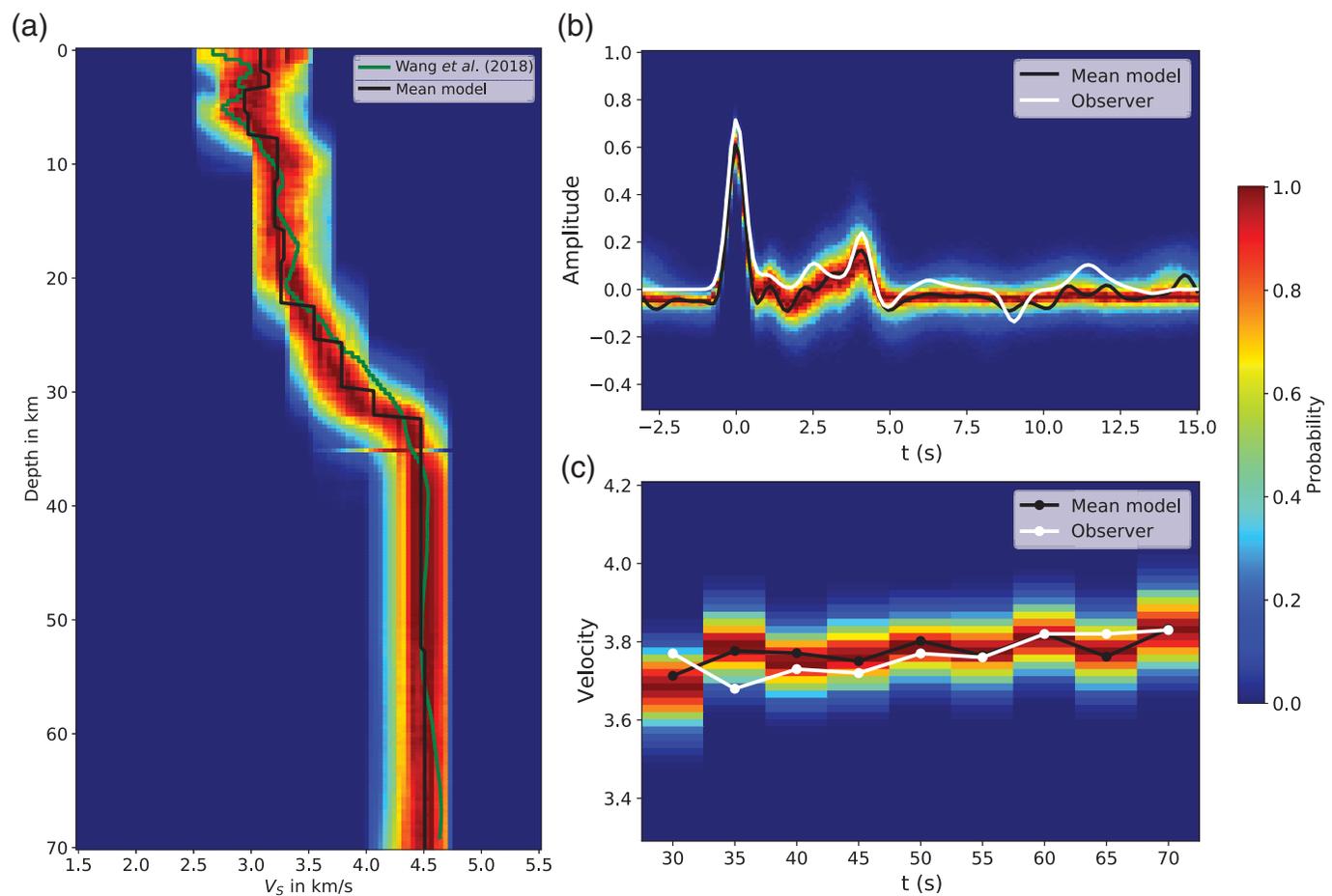


Figure A1. Inversion result of IU.CHTO with HMC framework. (a) Inversion result; (b) receiver function fitting; and (c) phase

velocity of Rayleigh wave fitting. The Gauss factor of receiver function is 2.5.

ALGORITHM

One Step of Hamiltonian Monte Carlo (HMC) Sampling

Input: Current sample x_{cur} ; time step length dt ; number of steps L ; boundary of samples x_{bnd} ; and Fréchet kernel ∇U

Output: New sample x_{new}

1. Initialize the momentum $P \sim N(0,1)$
2. Calculate the kinetic energy K and potential energy U
3. Compute current Hamiltonian $H_{cur} = K + U$
4. Update the momentum $P_{new} = P - dt \times \nabla U(x_{cur}) \times 0.5$
5. **For** $i = 1$ **to** L **do**
6. Update the current sample $x_{new} = x_{cur} + dt * P_{new}$
7. **If** x_{new} **is not inside** boundary
8. Adjust the current sample $x_{new} = 2 \times x_{bnd} - x_{new}$
9. **GOTO** Step 7
10. Update the momentum $P_{new} = P - dt \times \nabla U(x_{cur}) \times 0.5$
11. Calculate the new kinetic energy K_{new} and potential energy U_{new}
12. Compute new Hamiltonian $H_{new} = K_{new} + U_{new}$
13. **If** $\log[\text{random}(0, 1)] < \min(H_{cur} - H_{new}, 0)$
14. Accept the new sample: **Return** x_{new}
15. **Else**
16. Reject the new sample: **Return** x_{cur}

TABLE A1

Parameter for Real Data Inversion

Layers	V_s (km/s)	Thickness (km)
1st–4th	2.5–3	1–3
5th	2.5–3.7	2–5
6th–10th	3.2–3.7	
11st–13rd	3.5–4.5	
14th–16th	4.2–4.7	10–15
Half-space		\

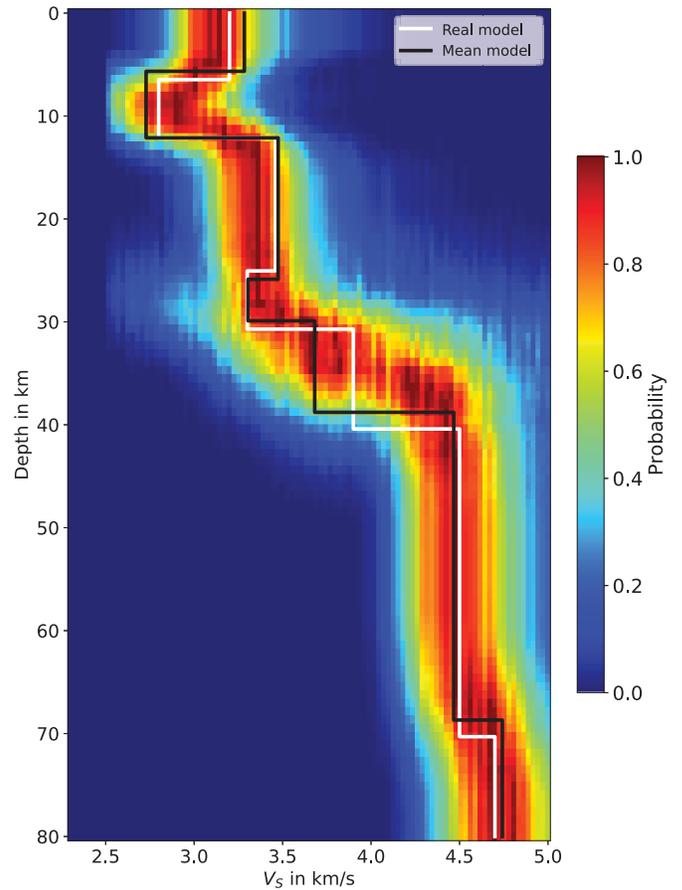


Figure A2. Inversion results of model 4 with wider search range.

Manuscript received 11 February 2022

Published online 14 October 2022