FULL PAPER

Comparison between the Hamiltonian Monte Carlo method and the Metropolis–Hastings method for coseismic fault model estimation

Taisuke Yamada[1*] , Keitaro Ohno[2] and Yusaku Ohta[1,3,4]

Abstract

A rapid source fault estimation and quantitative assessment of the uncertainty of the estimated model can elucidate the occurrence mechanism of earthquakes and inform disaster damage mitigation. The Bayesian statistical method that addresses the posterior distribution of unknowns using the Markov chain Monte Carlo (MCMC) method is significant for uncertainty assessment. The Metropolis–Hastings method, especially the Random walk Metropolis–Hastings (RWMH), has many applications, including coseismic fault estimation. However, RWMH exhibits a trade-off between the transition distance and the acceptance ratio of parameter transition candidates and requires a long mixing time, particularly in solving high-dimensional problems. This necessitates a more efficient Bayesian method. In this study, we developed a fault estimation algorithm using the Hamiltonian Monte Carlo (HMC) method, which is considered more efficient than the other MCMC method, but its applicability has not been sufficiently validated to estimate the coseismic fault for the first time. HMC can conduct sampling more intelligently with the gradient information of the posterior distribution. We applied our algorithm to the 2016 Kumamoto earthquake (M_{JMA} 7.3), and its sampling converged in 2×10^4 samples, including 1×10^3 burn-in samples. The estimated models satisfactorily accounted for the input data; the variance reduction was approximately 88%, and the estimated fault parameters and event magnitude were consistent with those reported in previous studies. HMC could acquire similar results using only 2% of the RWMH chains. Moreover, the power spectral density (PSD) of each model parameter's Markov chain showed this method exhibited a low correlation with the subsequent sample and a long transition distance between samples. These results indicate HMC has advantages in terms of chain length than RWMH, expecting a more efficient estimation for a high-dimensional problem that requires a long mixing time or a problem using nonlinear Green's function, which has a large computational cost.

Keywords: Hamiltonian Monte Carlo, Global navigation satellite system, Markov chain Monte Carlo, Real-time GEONET analysis system for rapid deformation, Bayesian inversion

Graphical Abstract

*Correspondence: taisuke.yamada.r3@dc.tohoku.ac.jp

[1] Research Center for Prediction of Earthquakes and Volcanic Eruptions,
Graduate School of Science, Tohoku University, 6-6 Aza-Aoba, Aramaki,
Aoba-ku, Sendai 980-8578, Japan
Full list of author information is available at the end of the article

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 2 of 16
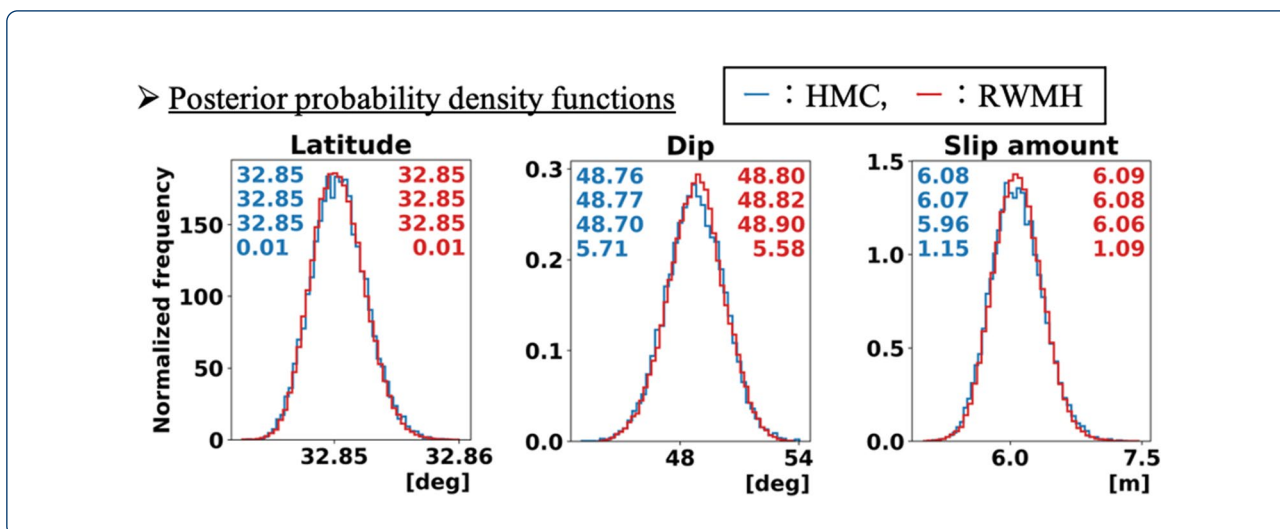
Introduction

To minimize earthquake and tsunami damage, the details of the coseismic fault must be estimated rapidly. Surface displacement data supplied by the global navigation satellite system (GNSS) can provide a stable estimation without underestimating the magnitude, considering that the measurement of a seismometer can be saturated when a megathrust occurs (Ohta et al. 2012). GNSS data are used in the real-time GEONET analysis system for rapid deformation monitoring (REGARD) (Kawamoto et al. 2017), which was jointly developed by the Geospatial Information Authority of Japan and Tohoku University. This system analyzes a GNSS carrier phase in real time to detect seismic events and automatically and rapidly estimate the fault model. Fault models were automatically estimated, for example, the 2016 Kumamoto earthquake (M_{JMA} 7.3) (Kawamoto et al. 2016), the 2019 Yamagata–Oki earthquake (M_{JMA} 6.7), and the 2021 Fukushima–Oki earthquake (M_{JMA} 7.3). These rapid inferences of REGARD include the slip distribution model on a plate interface in a subduction zone and a single rectangular fault estimation to address non-interplate earthquakes, such as inland and intra-slab earthquakes. The fault models estimated by REGARD were also adopted as part of the real-time tsunami inundation and damage forecast system operated by the Cabinet Office, Government of Japan (Musa et al. 2018; Ohta et al. 2018). It is important to accurately understand the estimation uncertainty of the estimated fault models to predict tsunamis and other natural disasters.

For uncertainty evaluation, a Bayesian statistics interpretation is often used to solve the inverse problem. This method is used to quantitatively assess the uncertainty of the prediction using the posterior probability density function (PDF) of model parameters while explicitly integrating prior information. Amey et al. (2018) applied the Bayesian inversion method to the inference of the slip distribution in the 2014 Napa Valley earthquake with self-similarity of the fault slip (Mai and Beroza 2002) as prior information. In addition, many previous studies have used Bayesian inversion (Sambridge and Mosegaard 2002; Fukuda and Johnson 2008; Dettmer et al. 2014; Ohno et al. 2022).

With non-interplate earthquakes such as the 2008 Iwate–Miyagi Nairiku earthquake (M_{JMA} 7.2) or the 2019 Yamagata–Oki earthquake (M_{JMA} 6.7), fault geometries are complex or unknown, and many earthquakes are difficult to assume the fault geometry in advance (Ohta et al. 2008; Ohno et al. 2021). The assumption of fault geometry required in the slip distribution model can make inferences ambiguous and affect the inferred result (Fukahata and Wright 2008; Fukuda and Johnson 2010; Duputel et al. 2014; Agata et al. 2021; Dutta et al. 2021).

Previous studies have satisfied the two requirements of rapid estimation and quantitative uncertainty assessment; for example, Minson et al. (2014) developed a real-time estimation method using Bayesian linear regression including the fault geometry selection with the assessment of its uncertainty and the slip distribution estimation on that geometry. Ohno et al. (2021) also developed Bayesian fault estimation methods that acquire the uncertainty of the fault parameters of a single rectangular fault by directly drawing the PDF using the Markov chain Monte Carlo (MCMC) method for real-time purpose. Their method incorporates the parallel tempering (Swendsen and Wang 1986; Geyer 1991) and an algorithm to automatically determine hyperparameters for real time and automatic fault estimation.

An MCMC sampling method typically used in the context of Bayesian fault estimation is the Random

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 3 of 16

walk Metropolis–Hastings (RWMH) method (Metropolis et al. 1953; Hastings 1970). RWMH forms a Markov chain based on the judgment of acceptance or rejection, depending on the ratio of the posterior distribution between a current parameter set and a transition candidate sampled randomly. Because of its simplicity, this algorithm has been applied to many studies in various fields, including fault estimation. However, RWMH exhibits a disadvantageous trade-off relationship between the acceptance ratio and the transition distance of parameters. RWMH sampling tend to decrease its acceptance ratio in proportion to the transition distance in the parameter space, especially for the sampling after the burn-in. Although the transition distance is often adjusted based on an acceptance ratio of approximately 30 to 50%, this low acceptance makes the inference inefficient. Furthermore, regarding the higher dimensional problems, the relevant volume shrinks in relation to the total model space volume (the curse of dimensionality), thus the acceptance ratio is low and the transition distance is further shortened. To address the trade-off problem, the RWMH method requires a relatively long chain and mixing time.

Considering this problem, we focus on the Hamiltonian Monte Carlo (HMC; Duane et al. 1987) method instead of the RWMH method. The HMC method requires a relatively short Markov chain to converge a sampling because this method can transition long distances while maintaining a higher acceptance ratio by utilizing the gradient information of the posterior PDF. Fichtner and Simute (2018) applied the weakly nonlinear problem of source point determination and developed an efficient and accurate algorithm. This study also proposed application to higher dimensional problems to capitalize on the efficiency of HMC–HMC is frequently used in the study of seismic tomography (Fichtner et al. 2019; Gebraad et al. 2020; Muir and Tkalčić 2020). However, HMC has not been applied to fault estimation problems in the past and its applicability needs to be verified.

In this study, as an application test, we developed an algorithm for a single rectangular fault estimation using HMC. To evaluate its performance, the algorithm was applied to the 2016 Kumamoto earthquake. Furthermore, to examine the accuracy and efficiency of the HMC method estimation, the results were compared with those of the conventional RWMH method. We mainly discuss the applicability of HMC to the estimation of earthquake source fault models and a total sample size required for our estimation, but only have a brief discussion of the computational time for real time usage.

Method
Bayes' theorem
Bayesian inversion is based on Bayes' theorem (Tarantola 2005; Gelman et al. 2021) is represented by Eq. (1) on a model parameter vector $\boldsymbol{\theta}$, and a data vector $\mathbf{d}$, used in estimation:

$$f(\boldsymbol{\theta}|\mathbf{d}) = \frac{f(\boldsymbol{\theta})f(\mathbf{d}|\boldsymbol{\theta})}{f(\mathbf{d})} \tag{1}$$

where $f(\boldsymbol{\theta}|\mathbf{d})$ is the posterior PDF, $f(\boldsymbol{\theta})$ is the prior PDF, $f(\mathbf{d}|\boldsymbol{\theta})$ is the likelihood function, and $f(\mathbf{d})$, which does not depend on the model parameters, is the normalization constant. According to Eq. (1), the posterior PDF is proportional to the prior PDF and the likelihood function.

Sampling method
Hamiltonian Monte Carlo method
We developed a method for single rectangular fault estimation using the HMC method (Duane et al. 1987; Neal 2011) and evaluated its ability in this study. First, in the HMC method, an auxiliary parameter vector $\boldsymbol{p}$ with the same dimensions as the model parameter vector $\boldsymbol{\theta}$, is introduced. Then, the Hamiltonian H is expressed by Eqs. (2) and (3) in the phase space at position $\boldsymbol{\theta}$ and momentum $\boldsymbol{p}$:

$$H(\boldsymbol{\theta}, \boldsymbol{p}) = h(\boldsymbol{\theta}) + \tfrac{1}{2}\boldsymbol{p}^{\mathrm{T}}\boldsymbol{p} \tag{2}$$

$$h(\boldsymbol{\theta}) = -\log\left(f(\boldsymbol{\theta}|\mathbf{d})\right) \tag{3}$$

where $h(\boldsymbol{\theta})$ is the potential energy as a function of position $\boldsymbol{\theta}$, and the Hamiltonian H is calculated using the set of $\boldsymbol{\theta}$ and $\boldsymbol{p}$. Subsequently, the transition in the phase space is performed while satisfying the conservation of energy, keeping the H value, with some error arising from numerical integration. In the HMC method, the leapfrog method is used for the simulation because it ensures volume-preservation, time reversibility, and detailed balance. The standard procedure for a single step of the leapfrog method is shown in Eqs. (4)–(6):

$$p_i\left(t + \tfrac{1}{2}\right) = p_i(t) - \frac{\mathrm{e}}{2}\frac{\partial h(\boldsymbol{\theta}(t))}{\partial \theta_i} \tag{4}$$

$$\theta_i(t+1) = \theta_i(t) + \mathrm{e}p_i\left(t + \tfrac{1}{2}\right) \tag{5}$$

$$p_i(t+1) = p_i\left(t + \tfrac{1}{2}\right) - \frac{\mathrm{e}}{2}\frac{\partial h(\boldsymbol{\theta}(t+1))}{\partial \theta_i} \tag{6}$$

where θ_i and p_i are the ith unknown and auxiliary parameters, respectively, e is the step size, t is the current step number, and L is the total number of step. Equations (4)–(6) are iterated from $t = 1$ to L, where L is the total step

Yamada *et al. Earth, Planets and Space* (2022) 74:86

Page 4 of 16

size of the leapfrog. We adopted an automatic differentiation in the Python/PyTorch module to calculate the derivative in Eqs. (4) and (6). Note that a gradient calculation may improve its efficiency with the analytical differentiation of Eq. (3). Furthermore, e and L are hyperparameters, and their settings can considerably affect an estimation efficiency for a general inverse problem. Subsequently, a numerical error, including the final parameters θ_L and p_L, is evaluated using the acceptance ratio p_{accept} :

$$p_{accept} = \min(1, r) \tag{7}$$

$$r = \exp\left(H\left(\boldsymbol{\theta}^{(\tau)}, \boldsymbol{p}^{(\tau)}\right) - H\left(\boldsymbol{\theta}^{(\tau')}, \boldsymbol{p}^{(\tau')}\right)\right) \tag{8}$$

where $\left(\boldsymbol{\theta}^{(\tau)}, \boldsymbol{p}^{(\tau)}\right)$ represents the current parameter ($t = 1$ in leapfrog), where τ is the current sample number and $\left(\boldsymbol{\theta}^{(\tau')}, \boldsymbol{p}^{(\tau')}\right)$ dose the candidate parameter ($t = L$). In Eq. (8), the Hamiltonian is theoretically preserved and only a small numerical error affects Eq. (8); therefore, a higher probability of acceptance than RWMH is realized.

The HMC method enables sampling that transitions a long distance in parameter space with a high acceptance ratio through the above procedure. This procedure is conducted as follows:

(i) An initial parameter $\left(\boldsymbol{\theta}^{(1)}\right)$, total sample size (T), and the hyperparameters e and L are set.

(ii) Random numbers are generated from a standard normal distribution and set to $\boldsymbol{p}^{(\tau)}$.

(iii) Transition from the current point $\left(\boldsymbol{\theta}^{(\tau)}, \boldsymbol{p}^{(\tau)}\right)$ to the candidate point $\left(\boldsymbol{\theta}^{(\tau')}, \boldsymbol{p}^{(\tau')}\right)$ is performed using the leapfrog method and hyperparameters e and L (Eqs. 4–6).

(iv) The candidate point is accepted or rejected using the acceptance ratio p_{accept}.

(v) (ii)–(iv) are iterated from $\tau = 1$ to T.

Figure 1 shows an example of exploration using the HMC method. As seen in the figure, HMC acts on the settings where the hyperparameters are $(e, L) = \left(10^{-3}, 10\right)$.

No-U-Turn sampler

The HMC method requires hyperparameters e and L which have a significant effect on the sampling efficiency. Therefore, optimization is needed to use this method efficiently. If L is too small, the transition distance between samples may be too short and more iteration is needed. However, if L is too large, the leapfrog will generate "U-Turn" trajectories; therefore, the final transition distance can become shorter than the middle. To address
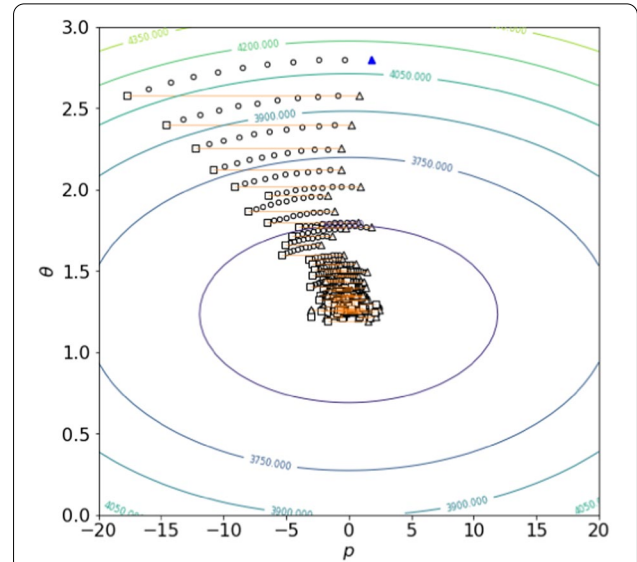

Fig. 1 Example of sampling using the Hamiltonian Monte Carlo (HMC) method. This figure shows the exploration of a mean parameter of a simple normal distribution from synthetic gaussian data $\left(f\left(\theta|d\right) \propto \exp\left(-\theta^2\right)\right)$. The contours show the constant Hamiltonian and their values. Triangles, circles, and squares show the start points, transition points, and end points of each leapfrog transition, respectively. The blue triangle particularly represents the initial point of the sampling. In addition, the orange line shows the jump to the next momentum generated by standard normal distribution after acceptance of the candidate point

these problems, especially to avoid the "U-Turn" transition, we used the No-U-Turn Sampler (NUTS; Hoffman and Gelman 2014), the extended method of the HMC.

Although the usual HMC uses the constant L throughout the sampling, the NUTS is designed not to need the constant L in sampling. Instead of stopping at the Lth time's integration, the leapfrog transition is terminated based on the detection of "U-Turn". The "U-Turn" in phase space defines the shortening of the distance Q between a leftmost point, $\boldsymbol{\theta}^+$, and a rightmost point, $\boldsymbol{\theta}^-$, of leapfrog transition, it is determined using the time derivative of Q:

$$Q = \tfrac{1}{2}\left(\boldsymbol{\theta}^+ - \boldsymbol{\theta}^-\right)^{\mathrm{T}}\left(\boldsymbol{\theta}^+ - \boldsymbol{\theta}^-\right) \tag{9}$$

$$\tfrac{dQ}{d\tau} = \left(\boldsymbol{\theta}^+ - \boldsymbol{\theta}^-\right)^{\mathrm{T}}\boldsymbol{p}^+ < 0 \text{ or } \left(\boldsymbol{\theta}^+ - \boldsymbol{\theta}^-\right)^{\mathrm{T}}\boldsymbol{p}^- < 0 \tag{10}$$

where the current sample number τ represents the time in the phase space, and $\boldsymbol{p}^+$ and $\boldsymbol{p}^-$ are the momentum at $\boldsymbol{\theta}^+$ and $\boldsymbol{\theta}^-$, respectively.

The above method enables the flexible escape of U-Turn, but the arbitrary stop of the leapfrog breaks the time reversibility. The NUTS algorithm overcomes this issue to introduce a procedure such as slice sampling, one

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 5 of 16

of the MCMC method. This is done by exploring from the current point $\boldsymbol{\theta}^{(\tau)}$ by building a binary tree to a random time direction (a sign of L), selecting the candidate state set C based on the condition below, and eventually selecting the subsequent states from C uniformly:

$$C = \{\boldsymbol{\theta}'|\boldsymbol{\theta}' \in B \cap H(\boldsymbol{\theta}',\boldsymbol{p}') < u\} \tag{11}$$

where B is the set of all visited states and u is the so-called slice generated as a uniform random number between 0 and the current posterior probability $\exp\left(-H\left(\boldsymbol{\theta}^{(\tau)},\boldsymbol{p}^{(\tau)}\right)\right)$.

Comparison with Metropolis–Hastings

To verify the accuracy and efficiency of our method, we used the Random walk Metropolis–Hastings (RWMH) that was adopted by Ohno et al. (2021) to compare the estimation by the HMC method to that by the RWMH method (Metropolis et al. 1953; Hastings 1970), which is often used to solve the geophysical problem (Fukuda and Johnson 2008; Ito et al. 2012; Dettmer et al. 2014; Minson et al. 2014; Ohno et al. 2022). The basic part of the RWMH we used, is detailed in Ohno et al. (2021). Although Ohno et al. (2021) introduced an automatic determination method for likelihood variance and the parallel tempering, we use the same variance value as HMC and estimate using only a single Markov chain for a single parameter to strictly compare both methods. Ohno et al. (2021) developed the method to automatically adjust the variances of the perturbation of parameters. However, here we adjusted it manually to transition quickly from the initial point of parameters on the sampling without the parallel tempering.

Setting of a single rectangular fault estimation

We estimated parameter $\boldsymbol{\theta}$ of the single rectangular fault model in Okada (1992) from permanent displacement data $\mathbf{d}$ based on real-time GNSS by the HMC and NUTS algorithms. The parameter vector $\boldsymbol{\theta}$ includes length and angle parameters of the rectangular in a uniform elastic half-space, that is, latitude and longitude of the fault plane, top depth, strike, dip, rake, fault length, fault width, and slip amount. In this section, we describe the setting of the estimation, such as the prior distribution or likelihood function.

Prior distribution

We prepared a normal distribution as prior distributions of latitude and longitude because we can use some external information such as the Earthquake Early Warning system by the Japan Meteorological Agency, and a uniform distribution such as a non-informative prior distribution of other parameters. The standard deviation of 2°

for latitude and longitude prior, is slightly large but has been set roughly for a weak restriction. In addition, we set a uniform distribution for the stress drop Δsd which is known to fall within a certain range. The upper and lower range of the prior is based on Ohno et al. (2021), which refers to the fault's scaling law. The stress drop is calculated using the following equation:

$$\Delta \mathrm{sd} = \frac{2c\mu S}{\sqrt{FL \cdot FW}} \tag{12}$$

where c and μ are a geometric factor and rigidity, respectively. We assumed these parameters as 0.5 and 30 [GPa], respectively. FL, FW, and S denote the length, width, and slip amount of the rectangular fault, respectively. We also established a uniform distribution for the ratio of fault width to fault length because the length is longer than the width usually. All prior distributions are shown in Table 1.

Likelihood function

The likelihood function is expressed by Eqs. (13) and (14):

$$f(\mathbf{d}|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{h}}^2}^{2N}} \exp\left(\frac{-1}{2\sigma_{\mathrm{h}}^2}\boldsymbol{r}_{\mathrm{h}}^{\mathrm{T}}\boldsymbol{r}_{\mathrm{h}}\right) \times \frac{1}{\sqrt{2\pi\sigma_{\mathrm{v}}^2}^{N}} \exp\left(\frac{-1}{2\sigma_{\mathrm{v}}^2}\boldsymbol{r}_{\mathrm{v}}^{\mathrm{T}}\boldsymbol{r}_{\mathrm{v}}\right) \tag{13}$$

$$\boldsymbol{r}_i = \widehat{\boldsymbol{d}}_i(\boldsymbol{\theta}) - \mathbf{d}_i \tag{14}$$

where N is the number of stations, and $\widehat{\boldsymbol{d}}_i(\boldsymbol{\theta})$ is the surface displacement of model $\boldsymbol{\theta}$ calculated by the method of Okada (1992). Subscripts h and v represent the components of horizontal and vertical displacements, respectively, and we assumed that they were independent.

Table 1 Prior parameter for fault parameters and certain calculated parameters

Parameter	Prior
Latitude (°)	$N(\mathrm{init}, 4)$
Longitude (°)	$N(\mathrm{init}, 4)$
Depth (km)	$U(0, \infty)$
Strike (°)	$U(0, 360)$
Dip (°)	$U(0, 90)$
Rake angle (°)	$U(-180, 180)$
Fault length (km)	$U(0, \infty)$
Fault width (km)	$U(0, \infty)$
Slip amount (m)	$U(0, \infty)$
Stress drop (MPa)	$U(0.2, 21.2)$
width/length (–)	$U(0, 1)$

A mean of normal distribution for latitude and longitude is assumed as their initial parameter. A reference point of latitude and longitude is the fault center, and depth is fault top. $N(\mu, \sigma^2)$ represents a normal distribution with mean μ, variance σ^2 and $U(a, b)$ represents a uniform distribution with a lower limit a and an upper limit b

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 6 of 16

Equation (13) assumes that the residuals, including the model and observational error, follow the normal distribution with mean 0 and variance σ_h^2 and σ_v^2.

Change in variables

HMC and NUTS use a gradient from the posterior PDF to the transition parameters. We observed the gradients of latitude and longitude to be notably large, except for seven parameters, resulting in the very low efficiency of the exploration. A potential method for correcting the gradients is the application of an adjustment hyperparameter e, which is the coefficient of the gradient, but the parameter e is common among the parameters; therefore, addressing our case is complicated by the differences among each gradient. Therefore, we applied a change of variables to the seven parameters to change the parameter scale. In addition, the sampling can be disturbed by a rejection due to a running to out of the parameter's domain, where 0 probability area, so we adopted log and logit translation as the change of variables to enlarge the sampling range. The depth D, fault length FL, fault width FW, and slip amount S were applied to the log translation (Eq. 15), and the strike $\emptyset$, dip δ, and rake angle λ were applied to logit (Eq. 16).

Log translation from $x \in (a, \infty)$ to $x' \in (-\infty, \infty)$:

$$x' = \log(x - a) \tag{15}$$

Logit translation from $x \in (a, b)$ to $x' \in (-\infty, \infty)$:

$$x' = \text{logit}(x) = \log\left(\frac{x-a}{b-x}\right) \tag{16}$$

Equations (15) and (16), for example, $D \in (0, \infty)[\text{km}]$ are translated to $D' = \log(D)$ and $\emptyset \in (0, 360)[°]$ to $\emptyset' = \log\{\emptyset/(360 - \emptyset)\}$. These changes of the sampling parameter implicitly assume not to include each value on the domain boundary of parameters, improve the efficiency of exploration, and expand practically the exploration range to the whole of real space. To correct our PDF using the Jacobian formula along with change variables, we estimated the following PDF by adding Jacobian **J** (Tarantola 2005; Gelman et al. 2021):

$$f(\boldsymbol{\theta}|\mathbf{d}) = \frac{1}{Z} \cdot f(\boldsymbol{\theta}) \cdot f(\mathbf{d}|\boldsymbol{\theta}) \cdot |\mathbf{J}| \tag{17}$$

$$|\mathbf{J}| = e^{D'} e^{FL'} e^{FW'} e^{S'} \frac{360 e^{\emptyset'}}{\left(1 + e^{\emptyset'}\right)^2} \frac{90 e^{\delta'}}{\left(1 + e^{\delta'}\right)^2} \frac{360 e^{\lambda'}}{\left(1 + e^{\lambda'}\right)^2} \tag{18}$$

where Z is a normalizing constant that is independent of $\boldsymbol{\theta}$. Note we set the parameter's domain as $\delta \in (0, 90)[°]$, $\lambda \in (-180, 180)[°]$, $FL \in (0, \infty)[\text{km}]$, $FW \in (0, \infty)[\text{km}]$, and $S \in (0, \infty)[\text{m}]$, respectively. It should be also described that the depth parameter is multiplied by a constant value before the log variable change to adapt to

the 2016 Kumamoto earthquake we discussed later. In this study, we chose 10^6 as the sufficiently large value, that means we treat depth as millimeter-scale, because the event occurred in a shallow place in the crust and actually the surface deformation was observed. This linear transformation of depth only involves multiplying the Jacobian (Eq. 18) by 10^{-6}. These log and logit translations in the posterior distribution were naturally reflected in the gradient computation detailed in "Hamiltonian Monte Carlo method" section.

Convergence test

In this study, the convergence of each Markov chain was evaluated using two criteria. First, we confirmed the smoothness of the histograms for each model parameter. Although it is a subjective perspective, it is required for convenience in later discussions using the central tendency value such as mode. Second, we used the method of convergence decision by Gelman (1996) for objective judgment. This method can determine the convergence by comparing a sample's variance between a full chain and a divided one. We applied this test to all model parameters individually and confirmed their convergence.

If the Markov chain whose length is T excluded burn-in, it is divided into K equal chains, and T/K length chains are prepared. Then the following two indices of a chain's variance are calculated, where $\bar{\theta}$ and $\bar{\theta}_k (k = 1, 2, ..., K)$ are the mean values of a parameter in the full chain and the kth divided chain, respectively:

$$\text{VAR}_B = \frac{T}{K-1} \sum_{k=1}^{K} \left(\bar{\theta}_k - \bar{\theta}\right)^2 \tag{19}$$

$$\text{VAR}_W = \frac{1}{K(T-1)} \sum_{k=1}^{K} \sum_{t=1}^{T} \left(\theta_k^{(t)} - \bar{\theta}_k\right)^2 \tag{20}$$

Equations (19) and (20) roughly represent the variance between the chains and the average variance of each chain, respectively. Then, the index of a convergence decision using the two indices above is expressed in the following equation:

$$R = \sqrt{\frac{T-1}{T} + \frac{1}{T} \frac{\text{VAR}_B}{\text{VAR}_W}} \tag{21}$$

The R value is almost 1 on the converged chain; in fact, limitless sampling must yield a converged Markov chain; it is explicitly stated for Eq. (21) that the R value converges to 1, where $T \to \infty$. In this study, the convergence threshold is given as $R < 1.1$, according to Gelman (1996).

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 7 of 16

Data and the setting of hyperparameters

We applied our method to the 2016 Kumamoto earthquake (M_{JMA} 7.3), which occurred at 16:25:05 (UTC) in the Kyusyu region, southwestern Japan. Ground displacement data including three components (east–west, north–south, and up-down) were obtained using the REGARD system (Kawamoto et al. 2016) in real time, and we used the data at 60 s after the mainshock at 200 stations near the source (Fig. 2).

Table 2 shows the settings of the hyperparameters in both methods, HMC and RWMH estimation. The total

Table 2 Setting of hyperparameters

	Leapfrog's step size e	Variance in likelihood (cm²)	Chain length (count)	Burn-in chain (count)
HMC	1×10^{-3}	4	2×10^4	1×10^3
RWMH	–	4	1×10^6	5×10^4

number of step, L, was optimized in each leapfrog step using the NUTS algorithm explained above. In contrast, the step size of leapfrog e was determined manually. The
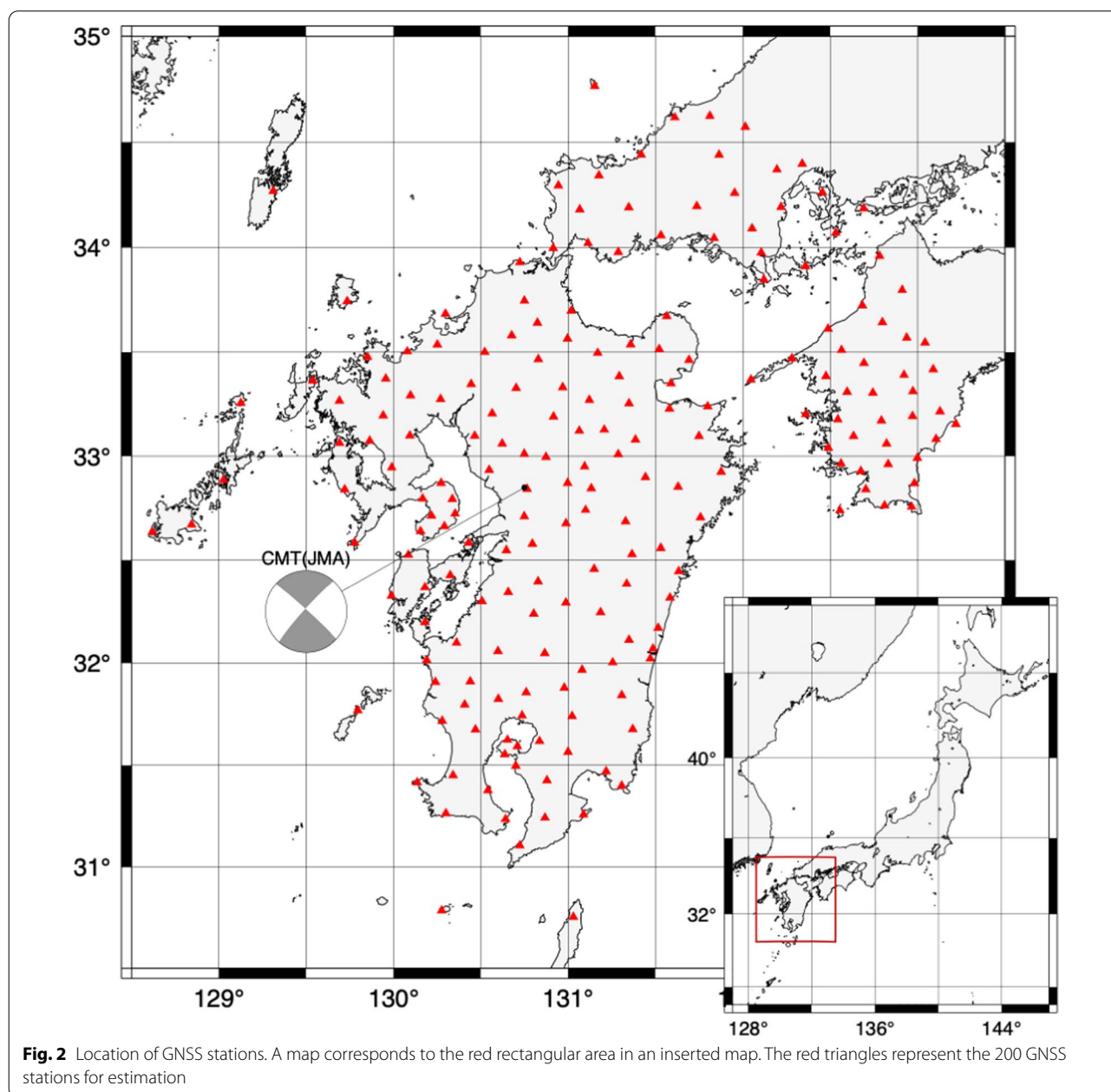
Fig. 2 Location of GNSS stations. A map corresponds to the red rectangular area in an inserted map. The red triangles represent the 200 GNSS stations for estimation

Yamada *et al. Earth, Planets and Space* (2022) 74:86

Page 8 of 16

step size was adjusted a priori by halving or doubling the attempt according to the initial VR's soaring. There is still a hyperparameter mass matrix to be optimized, which is a covariance matrix of the normal distribution for momentum $\boldsymbol{p}$ (that means to use the normal distribution with arbitrary covariance as the proposal distribution of $\boldsymbol{p}$) and relates with a mixing time of sampling. Typically, the mass matrix is optimized adaptively in a burn-in chain and the efficiency of estimation can be ameliorated. However, in this study, we have not adjusted the hyperparameter to simplify the algorithm and leave it for future work. The length of the entire chain was determined using the convergence test explained above, and the length of the burn-in chain was determined by evaluating the trace plot of parameters, especially variance reduction (VR) calculated by the following equation:

$$\text{VR} = 100 \times \left(1 - \frac{\boldsymbol{r}^{\mathrm{T}} \boldsymbol{r}}{\mathbf{d}^{\mathrm{T}} \mathbf{d}} \right) \tag{22}$$

where $\mathbf{d}$ and $\boldsymbol{r}$ denote the vector of all the displacement data and residuals between the data and model displacement ($\mathbf{d} = \left(\mathbf{d}_{\mathrm{h}}^{\mathrm{T}}, \mathbf{d}_{\mathrm{v}}^{\mathrm{T}} \right)^{\mathrm{T}}$ and $\boldsymbol{r} = \left(\boldsymbol{r}_{\mathrm{h}}^{\mathrm{T}}, \boldsymbol{r}_{\mathrm{v}}^{\mathrm{T}} \right)^{\mathrm{T}}$), respectively.

Additional file 1: Figures S1 and S2 show the results of the convergence test and the entire trace plot of each parameter, respectively. We continued the HMC sampling with 20,000 samples because the parameters should satisfy both criteria; nevertheless, these figures indicate the convergence with 3000 samples. The burn-in samples are often determined using a trace plot of VR or posterior probability (Amey et al. 2018); therefore, we also removed the first 5% of the chain similarly. Although the observation error of GNSS may appear to result from the horizontal error being smaller than the vertical error, we assumed the standard deviation in likelihood, σ_{h} and σ_{v}, representing the same error value, $\sigma_{\mathrm{h}} = \sigma_{\mathrm{v}} = 2$ cm to simplify the inversion.

Table 2 also shows the hyperparameters in terms of RWMH sampling. The entire burn-in chain length was determined the same way as the HMC, a factor of 50, compared to HMC (Additional file 1: Figs. S1, S2).

Results

Expression of posterior probability density functions

For geophysical interpretation, posterior PDFs regarding the fault parameters are required. However, we can directly obtain the rescaled PDFs because of the change in variables (detailed in "Change in variables" section). Moreover, the changes using log and logit function whose Jacobian is not identity matrix complicate the restoration of their scale. Therefore, we developed a special plotting for log and logit translation. We firstly made the histograms on a changed parameter scale and inversely

changed the upper and lower boundary values of class division while maintaining the class frequency. Then, we plotted the normalized histograms on original parameter's scale using the boundary values. The change in appearance following the use of our expression method is shown in Additional file 1: Fig. S3.
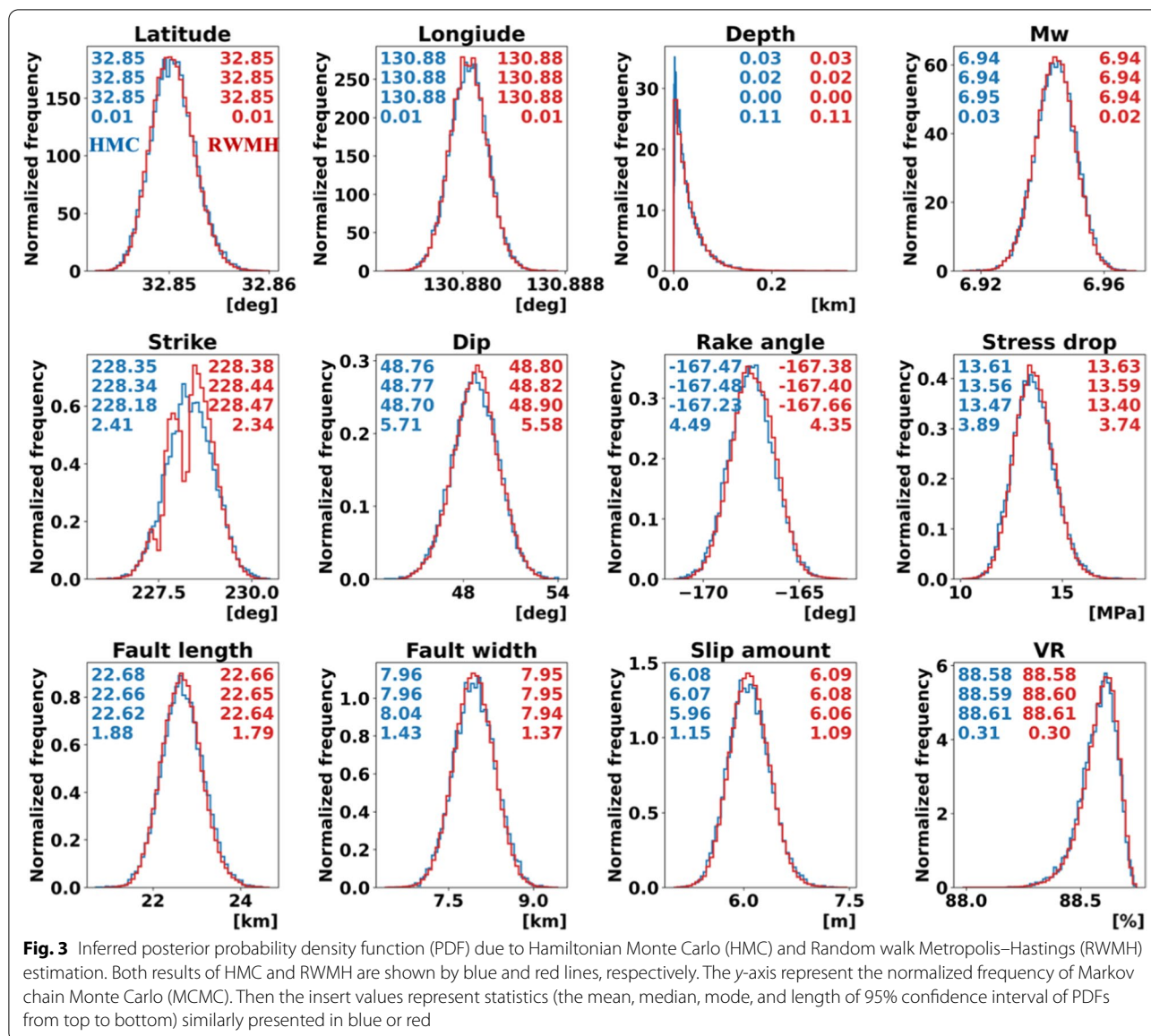
Results of the application for the Kumamoto earthquake

Figure 3 shows the posterior PDFs of each unknown parameter (latitude, longitude, depth, strike, dip, rake angle, fault length, fault width, and slip amount) and certain seismic parameters [moment magnitude (Mw), stress drop, and VR] computed for each sample inferred by HMC and RWMH.

The HMC results in Fig. 3 denote a shallow right-lateral fault dipping toward the northwest, and the displacement predictions calculated from each sample are consistent with the data (VR ∼ 88%). The distribution of Mw is a reasonable estimate with a peak of approximately 7.0, which is consistent with previous studies (Asano and Iwata 2016; Kawamoto et al. 2016; Yarai et al. 2016; Tanaka et al. 2019).

Figure 3 shows considerable features that the PDFs obtained using HMC agree with RWMH in the shape of PDFs or representative values (mean, median, and mode); nevertheless, both methods estimate independently. This means that HMC can be applied to the single rectangular fault estimation. Conversely, there are a few differences in the PDF of strike that RWMH's has multiple peaks. Although we conducted the RWMH and HMC sampling, whose chain length is 10 times the main result in the manuscript, they remain as they were (Additional file 1: Fig. S4). Both methods may shape different PDFs attributed to the difference between both sampling methodologies and to the implicit assumption of a positive depth on HMC sampling. However, the strike posterior distributes in an extremely narrow range ($\sim 4°$), which can only have a little effect on the surface deformation model. Therefore, we concluded their distributions were in rough agreement and will validate it in future by applying them to other events. There are also a few differences in PDFs of the depth. The outstanding peak of the depth's PDF of the HMC near 0 must be induced by the log variable changes, even if we deal with the millimeter depth. The log change can adjust a gradient of a posterior at almost entirely a domain, but a trend of the gradient changes in $x < 1$, resulting in that a transition distance of sampling is shortened and x less than 1 is often sampled. This is just an issue derived from the parametrization, so we would like to adopt a more effective one in the future.

For a comparison of surface deformation, Fig. 4a shows the rectangular fault of a mode model (the so-called maximum a posteriori model) solved by HMC,
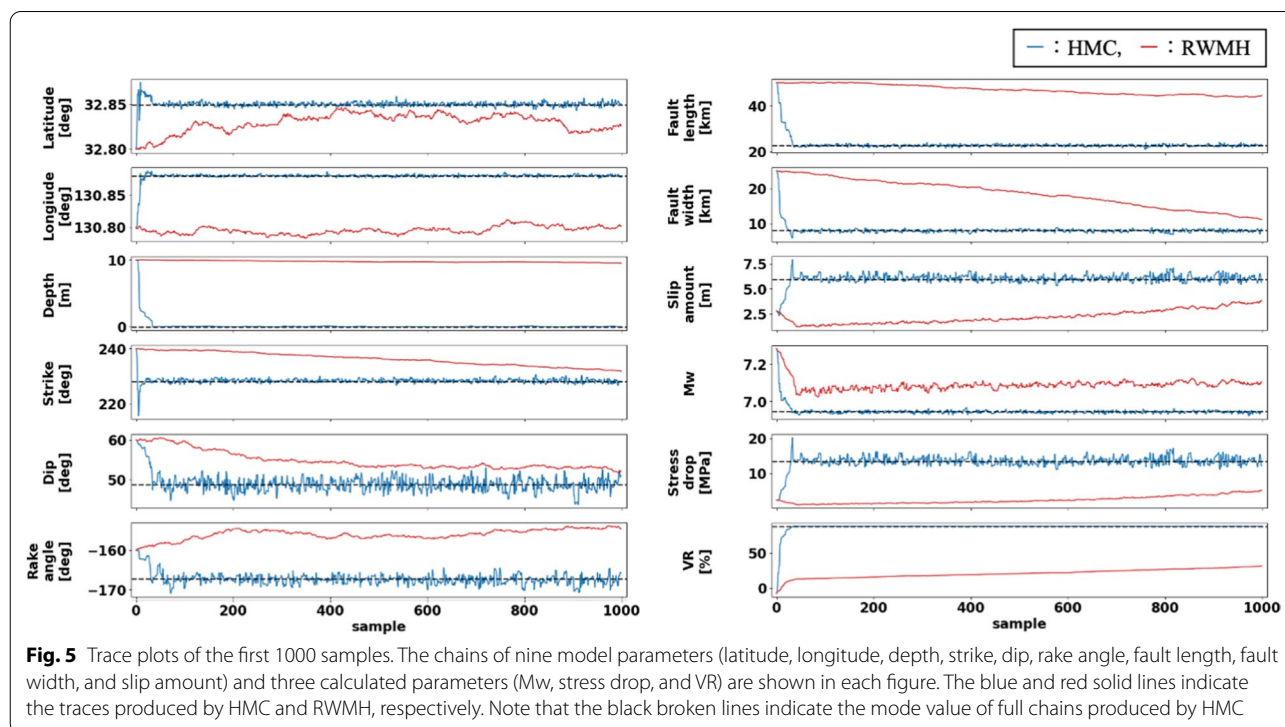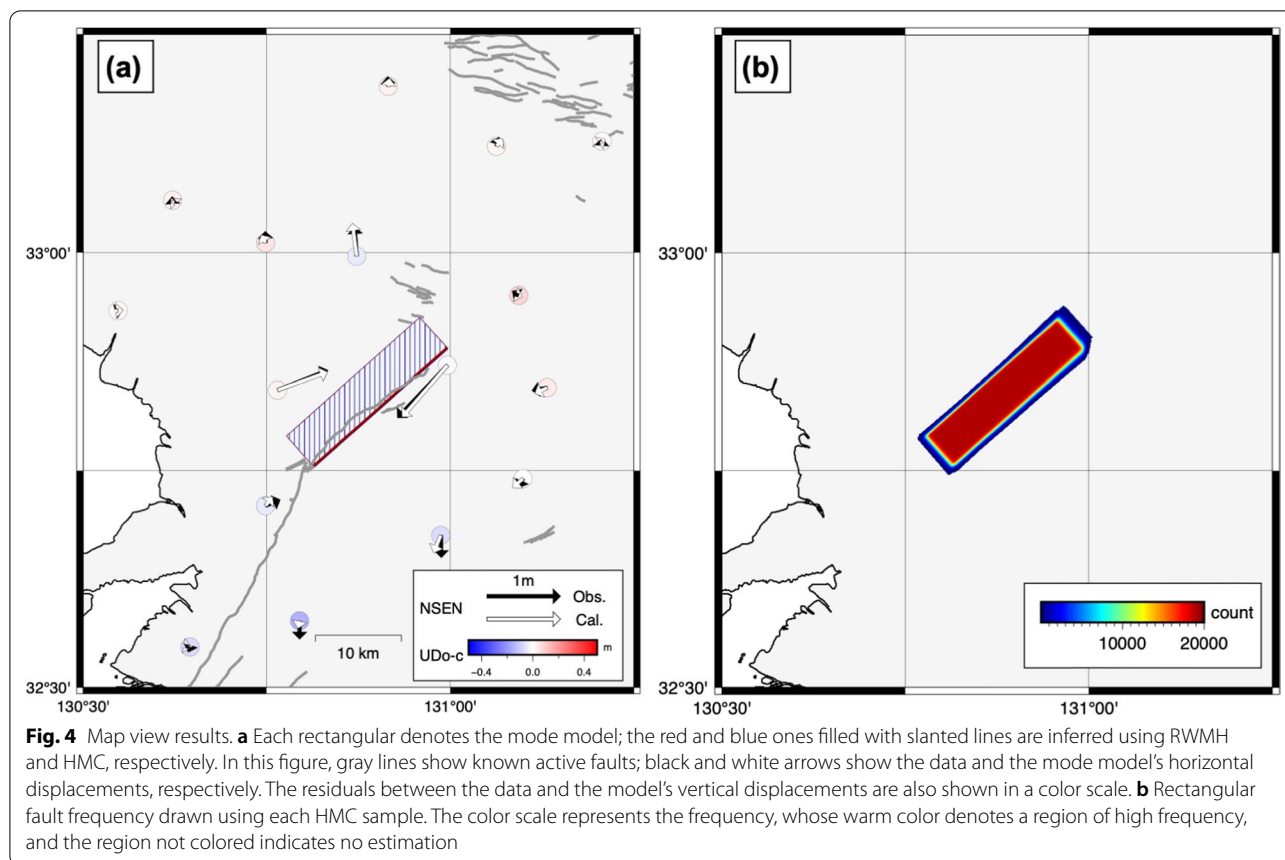
Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 9 of 16

Fig. 3 Inferred posterior probability density function (PDF) due to Hamiltonian Monte Carlo (HMC) and Random walk Metropolis–Hastings (RWMH) estimation. Both results of HMC and RWMH are shown by blue and red lines, respectively. The *y*-axis represent the normalized frequency of Markov chain Monte Carlo (MCMC). Then the insert values represent statistics (the mean, median, mode, and length of 95% confidence interval of PDFs from top to bottom) similarly presented in blue or red

and the predicted displacement at each station assumed the model with the displacement data we used. Figure 4 also shows a mode model solved by RWMH. As already mentioned, the HMC mode model adequately explains the displacement data and is consistent with the RWMH model. As illustrated in Fig. 4b, which is drawn using a method proposed by Ohno et al. (2021), we exhibited the spatial frequency of the fault rectangular. This figure includes the color scale, considering overlaps among the rectangles of each HMC sample on the map. HMC models likely contain only a small uncertainty of estimation because the color scale expands in a significantly narrow region.

Discussion
Comparison of initial exploration
Figure 5 shows the Markov chains of the first 1000 samples in the HMC and RWMH methods. The HMC exploring can produce estimates approximately the same as the final products within 40 samples and VR of 88%. In contrast, the trace plots of RWMH appear that the method is still seeking a more optimized solution at even 1,000 samples, and then approximately 20,000 samples were spent to improve VR to 88% (Additional file 1: Fig. S2). Although now we compared both the methods in only a single case, that results are important to ensure efficient convergence of HMC, where HMC required less than 1% chains of RWMH's for model exploration. To use this

Yamada *et al. Earth, Planets and Space* (2022) 74:86

Page 10 of 16

Fig. 4 Map view results. **a** Each rectangular denotes the mode model; the red and blue ones filled with slanted lines are inferred using RWMH and HMC, respectively. In this figure, gray lines show known active faults; black and white arrows show the data and the mode model's horizontal displacements, respectively. The residuals between the data and the model's vertical displacements are also shown in a color scale. **b** Rectangular fault frequency drawn using each HMC sample. The color scale represents the frequency, whose warm color denotes a region of high frequency, and the region not colored indicates no estimation

Fig. 5 Trace plots of the first 1000 samples. The chains of nine model parameters (latitude, longitude, depth, strike, dip, rake angle, fault length, fault width, and slip amount) and three calculated parameters (Mw, stress drop, and VR) are shown in each figure. The blue and red solid lines indicate the traces produced by HMC and RWMH, respectively. Note that the black broken lines indicate the mode value of full chains produced by HMC

efficiency, this method should be applied to high-dimensional problems, such as slip distribution estimation or inversion using nonlinear Green's function, which incurs a high computational cost for the forward calculation.

The motion of exploration appearing in trace plots also denotes an obviously high efficiency. The rake angle or slip amount chains using the RWMH method initially move in the opposite direction of the mode value of whole sampling, but the chains using the HMC method directly select the plausible direction. This shows that the HMC method is superior to the RWMH method because of the transition using gradients of the posterior PDF. Chains of latitude or strike denote that, even if HMC passes by following the mode value, a quick return to the same value is expected. It is an interesting exploration tendency showing the ability of the HMC method to explore a wide range of parameters in a single transition and aim at a minimum posterior immediately. These features cannot be realized by the RWMH method, whose transition is performed by adding only a small amount and whose acceptance ratio is generally low.

Figure 6 shows the marginal PDFs that appear in pairs of parameters, fault length, fault width, and slip amount. Figures using all parameters are also shown in Additional file 1: Figs. S5–S8. Figure 6a (HMC) and Fig. 6b (RWMH) display the marginal PDFs including the burn-in samples, indicating that the HMC method can reach the high probability area effectively with straight and stepping trace. This method only traces the plausible area because of the larger white area which has never been explored and smaller blue area whose samples are low posterior probability rather than the RWMH method. Considering that Fig. 6c (HMC) and Fig. 6d (RWMH) show that the HMC method can generate a similar correlation to RWMH among parameters, and we confirmed HMC can grasp the correlation more efficiently.

However, the HMC or other MCMC methods have the problem of ensuring minimum global exploration. One method to explore more broadly, with certainty, and efficiently in parameter space is the parallel tempering (Dosso et al. 2012; Sambridge 2013; Dettmer et al. 2014; Hallo and Gallovič 2020; Ohno et al. 2021, 2022). In future work, we will develop a method that introduces this expanding method to the HMC and realize a more efficient and stable fault estimation.

Comparison regarding the autocorrelation of Markov chains

The RWMH method conducts the sampling of parameter exploration by slightly perturbating the model parameters. In contrast, HMC transitions from place to place and visits plausible samples if they are far away. Given these sampling characteristics, we analyzed the autocorrelation of Markov chains generated by both methods.

Figure 7 shows the power spectral densities of the Markov chain regarded as a time series and its inclination ("*a*" value in the figure) using log approximation. In this figure, the results of the RWMH method explicitly denote the inverse proportion to frequency f [/count] and show random walk tendency; the inclination is roughly f^{-2}. The inclination in the zone over 1000 samples appears to be flat, similar to white noise. However, the results of the HMC method denote an inclination of less than 0.5 and a white noise-like tendency for most parameters. Only the result of the depth parameter denotes an inclination larger than the HMC's other results, nearly 1.5, although it is slightly smaller than the RWMH method. This high autocorrelation is considered caused by exploration near the domain. Hallo and Gallovič (2020) address similar problems using the unique prior PDF reflected at the parameter domain.

For MCMC sampling, a higher autocorrelation induces deterioration of sampling efficiency and complicates the achievement of the proper posterior PDFs by finite-length chains (Geyer 2011). RWMH sampling is often highly correlated between close samples. The thinning method, which discards all but every kth sampled value, is likely to address this problem (Link and Eaton 2012), but it does not guarantee improved estimation efficiency (Geyer 2011; Link and Eaton 2012). In addition, even if we adopt this method to introduce RWMH, we must abandon several thousand samples for one sample because the PSD analysis shows a high autocorrelation in the zone over 1000 samples. Therefore, it is conceivable that the HMC method can efficiently achieve a lower autocorrelation sampling without wasting samples.

More optimized utilization of the HMC

At the point of the number of samples, compared to the RWMH method, the HMC method can estimate posterior PDFs more efficiently in the Bayesian analysis of coseismic fault estimation. However, the HMC can also be less efficient at the point of the calculation time, especially when applying to a problem where one sample's computational cost of the forward calculation is low, such as in the problem in this study. Additional file 1: Figure S9 shows the ratio of calculation time in the HMC algorithm. It shows that the backward calculation is more expensive than the forward in this problem, which is the inversion of the nine fault parameters. Therefore, the HMC can be used for problems that require copious amounts of resources for forward computation. We should also pay attention to the gradient calculation. The automatic differentiation we adopted to realize the complicated gradient calculation of Eq. (3) has to execute the
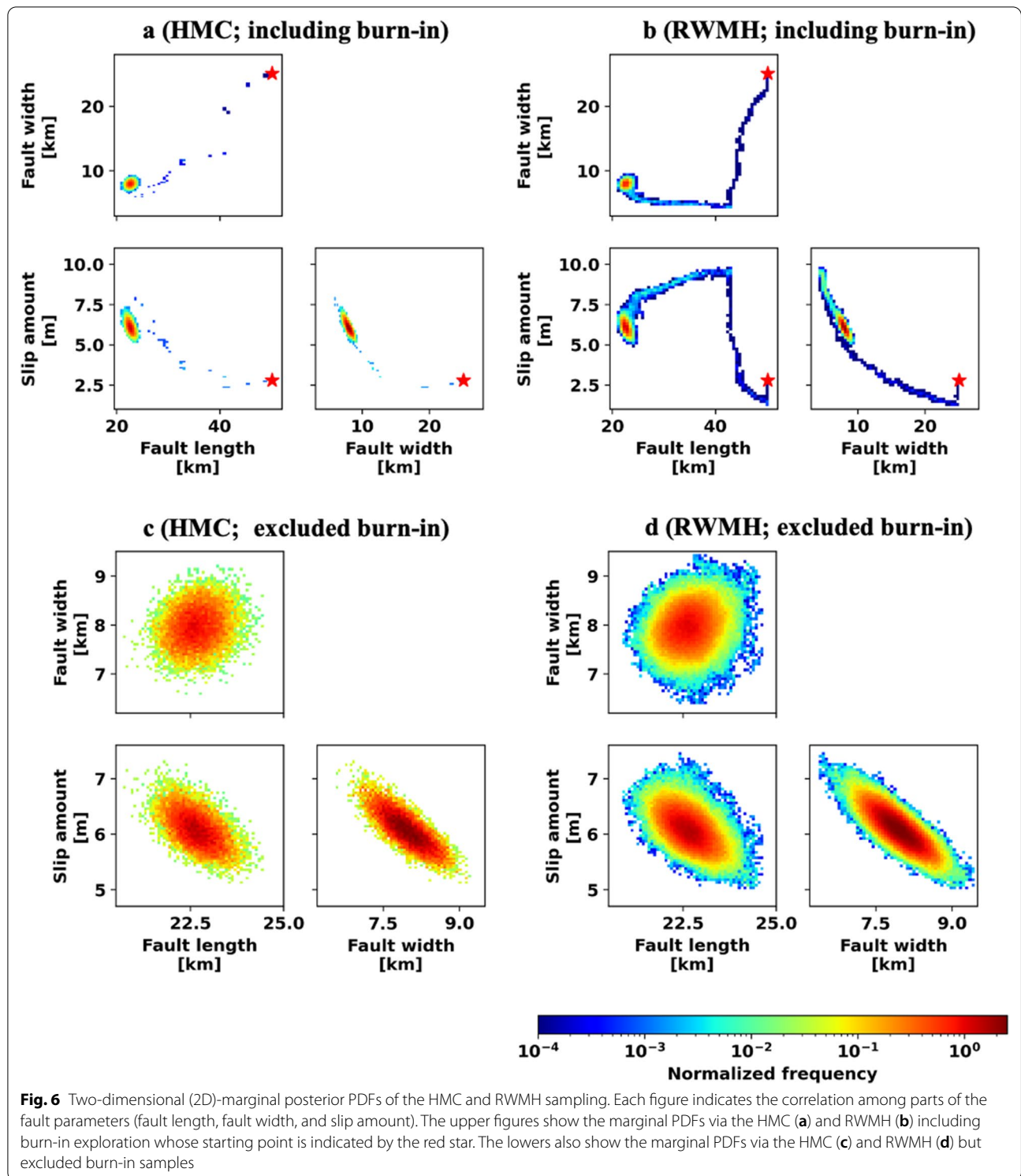
Fig. 6 Two-dimensional (2D)-marginal posterior PDFs of the HMC and RWMH sampling. Each figure indicates the correlation among parts of the fault parameters (fault length, fault width, and slip amount). The upper figures show the marginal PDFs via the HMC (**a**) and RWMH (**b**) including burn-in exploration whose starting point is indicated by the red star. The lowers also show the marginal PDFs via the HMC (**c**) and RWMH (**d**) but excluded burn-in samples

forward calculation previously. Additional file 1: Figure S9 shows that around 66% of forwarding is executed for the purpose of automatic differentiation. Thus, refining the gradient calculation method or using the analytic differential must improve the algorithm efficiency drastically.

To reduce the calculation time, a method for optimizing hyperparameter e, which represents the step size of the leapfrog, should be also developed. For example,

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 13 of 16
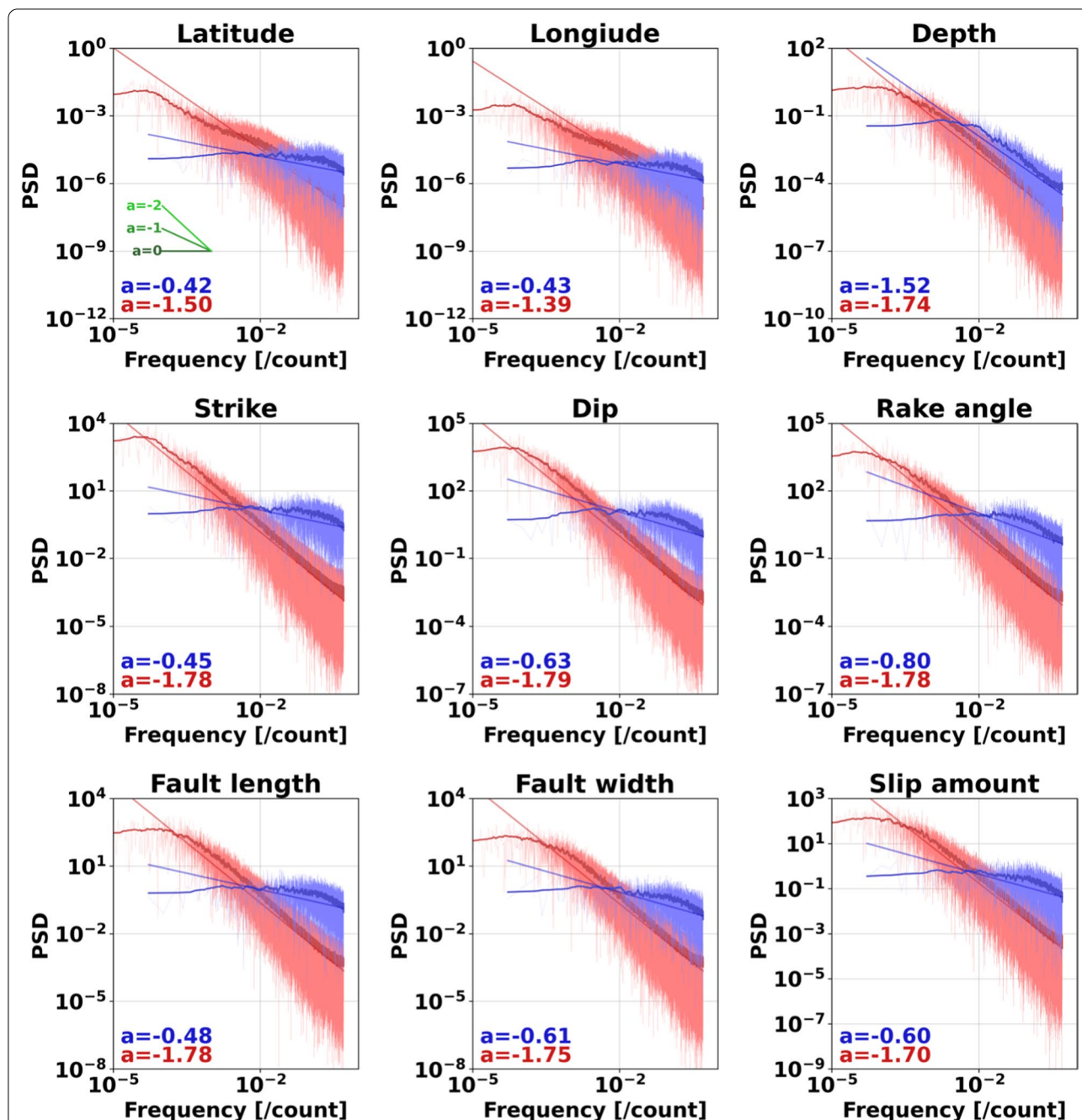
Fig. 7 Power spectral density (PSD) of Markov chains of nine model parameters. In each figure, red and blue show the results of RWMH and HMC, respectively. The light color line shows PDF and the heavy color line shows the moving average of PDF (the average of 25 points on the right and left side). The result of the log approximation is visually shown by the straight line and its estimated inclination is inserted on the bottom left of each figure (upper blue: HMC, lower red: RWMH)

Stan (Hoffman and Gelman 2014) employs the dual averaging method developed by Nesterov (2009), a stochastic gradient descent method, which is expected to optimize in the same way as our method.

Besides the hyperparameter, the calculation time of the estimation is based on the number of model parameters. It is difficult for the RWMH method to conduct efficient sampling in a parameter space that comprises many

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 14 of 16

parameters because the space is vast for model exploration (Roberts and Rosenthal 1998; Ringer et al. 2021). However, as reported by Fichtner and Simute (2018), the HMC method can maintain the calculation time by using the differentiation of the posterior PDF which contributes to ensuring the transition to the high posterior probability region. Therefore, the HMC can be superior and play an important role in solving high-dimensional problems, such as slip distribution estimation.

Conclusions

We developed a fault estimation algorithm using the HMC method; this was the first attempt to use the HMC method for fault estimation. The developed method performs efficiently following the adjustment of each parameter scale by changing the variables, well using the log and logit functions. We then applied the developed method to the 2016 Kumamoto earthquake (M_{JMA} 7.3) and assessed its feasibility for the single rectangular fault estimation and its exploration efficiency by comparison with the RWMH method.

We found that the HMC method estimates the informative model that reproduces the observation data and substantially corresponds to the fault model predicted by the RWMH method, although both algorithms perform the estimation independently. This indicates that the HMC method can be used for the single rectangular fault estimation. The length of the Markov chain required for HMC sampling was approximately 2% of that of the RWMH method, and 1% samples were needed to explore the high VR fault models (VR ~ 88%). We also clarified that the HMC method works as a low autocorrelation and a long transition from sample to sample, as described by the PSD analysis of the Markov chains of each parameter. These results are attributed to the theoretical characteristics of the HMC method, which allows for an efficient search following the gradient of the posterior distribution. This reasonable HMC sampling appears to be valid for high-dimensional problems, such as slip distribution estimation or problems using the nonlinear Green's function, which incurs high computational costs for one sample.

In contrast, the HMC method requires a longer calculation time for one sample than the RWMH method. This indicates that using the RWMH method is more suitable than that of the HMC for the estimation of a single rectangular fault, a low-dimensional problem. Therefore, the choice between the RWMH or HMC methods depends on the cost of forwarding calculation and the number of unknown parameters in the applied problem.

Abbreviations

HMC: Hamiltonian Monte Carlo; RWMH: Random walk Metropolis–Hastings; PDF: Probability density function; MCMC: Markov Chain Monte Carlo; REGARD: REal-time GEONET analysis system for rapid deformation monitoring; GNSS: Global Navigation Satellite System; VR: Variance reduction; PSD: Power spectral density.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40623-022-01645-y.

Additional file 1: Figure S1. Results of the convergence test. **Figure S2.** Trace plots of entire chains. **Figure S3.** Example of the expression of posterior PDF for variable changed parameters. **Figure S4.** Result when the chain length is 10 times the main result. **Figure S5.** Two dimensional (2D)-marginal posterior PDFs of the HMC sampling, including the Burn-in phase. **Figure S6.** 2D-marginal posterior PDFs of the RWMH sampling, including the burn-in phase. **Figure S7.** 2D-marginal posterior PDFs of the HMC sampling excluding the Burn-in phase. **Figure S8.** 2D-marginal posterior PDFs of the RWMH sampling excluding the burn-in phase. **Figure S9.** Pie chart of the calculation time HMC spend.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Research Center for Prediction of Earthquakes and Volcanic Eruptions, Graduate School of Science, Tohoku University, 6-6 Aza-Aoba, Aramaki, Aoba-ku, Sendai 980-8578, Japan. [2]Geospatial Information Authority of Japan, 1 Kitasato, Tsukuba 305-0811, Japan. [3]Division for the Establishment of Frontier Sciences of Organization for Advanced Studies, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan. [4]International Research Institute of Disaster

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 15 of 16

Science, Tohoku University, Aoba 468-1, Aramaki, Aoba-ku, Sendai 980-8572, Japan.

References

Agata R, Kasahara A, Yagi Y (2021) A Bayesian inference framework for fault slip distributions based on ensemble modelling of the uncertainty of underground structure: with a focus on uncertain fault dip. Geophys J Int 225:1392–1411. https://doi.org/10.1093/gji/ggab033

Amey R, Hooper A, Walters R (2018) Bayesian method for incorporating self-similarity into earthquake slip inversions. J Geophys Res Solid Earth 123:6052–6071. https://doi.org/10.1029/2017JB015316

Asano K, Iwata T (2016) Source rupture processes of the foreshock and mainshock in the 2016 Kumamoto earthquake sequence estimated from the kinematic waveform inversion of strong motion data. Earth Planets Space 68:1–11. https://doi.org/10.1186/s40623-016-0519-9

Dettmer J, Benavente R, Cummins PR, Sambridge M (2014) Trans-dimensional finite-fault inversion. Geophys J Int 199:735–751. https://doi.org/10.1093/gji/ggu280

Dosso SE, Holland CW, Sambridge M (2012) Parallel tempering for strongly nonlinear geoacoustic inversion. J Acoust Soc Am 132:3030–3040. https://doi.org/10.1121/1.4757639

Duane S, Kennedy AD, Pendleton BJ, Roweth D (1987) Hybrid Monte Carlo. Phys Lett B 195:216–222. https://doi.org/10.1016/0370-2693(87)91197-X

Duputel Z, Agram PS, Simons M, Minson SE, Beck JL (2014) Accounting for prediction uncertainty when inferring subsurface fault slip. Geophys J Int 197:464–482. https://doi.org/10.1093/gji/ggt517

Dutta R, Jónsson S, Vasyura-Bathke H (2021) Simultaneous Bayesian estimation of non-planar fault geometry and spatially-variable slip. J Geophys Res Solid Earth 126:1–28. https://doi.org/10.1029/2020JB020441

Fichtner A, Simutė S (2018) Hamiltonian Monte Carlo inversion of seismic sources in complex media. J Geophys Res Solid Earth 123:2984–2999. https://doi.org/10.1002/2017JB015249

Fichtner A, Zunino A, Gebraad L (2019) Hamiltonian Monte Carlo solution of tomographic inverse problems. Geophys J Int 216:1344–1363. https://doi.org/10.1093/gji/ggy496

Fukahata Y, Wright TJ (2008) A non-linear geodetic data inversion using ABIC for slip distribution on a fault with an unknown dip angle. Geophys J Int 173:353–364. https://doi.org/10.1111/j.1365-246X.2007.03713.x

Fukuda J, Johnson K (2008) A fully Bayesian inversion for spatial distribution of fault slip with objective smoothing. Bull Seismol Soc Am 98:1128–1146. https://doi.org/10.1785/0120070194

Fukuda J, Johnson K (2010) Mixed linear-non-linear inversion of crustal deformation data: Bayesian inference of model, weighting and regularization parameters. Geophys J Int 181:1441–1458. https://doi.org/10.1111/j.1365-246X.2010.04564.x

Gebraad L, Boehm C, Fichtner A (2020) Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo. J Geophys Res Solid Earth 125:1–18. https://doi.org/10.1029/2019JB018428

Gelman A (1996) Inference and monitoring convergence. Markov chain Monte Carlo in practice. Chapman & Hall/CRC, London

Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D (2021) Bayesian data analysis third edition (with errors fixed as of 15 February 2021). Chapman & Hall/CRC, London

Geyer CJ (2011) Introduction to Markov chain Monte Carlo. Handbook of Markov chain Monte Carlo. Chapman & Hall/CRC, London

Geyer CJ (1991) Markov chain Monte Carlo maximum likelihood. In: Proceedings of the 23rd symposium on the interface, computing science and statistics, pp 156–163

Hallo M, Gallovič F (2020) Bayesian self-adapting fault slip inversion with green's functions uncertainty and application on the 2016 Mw7.1 Kumamoto earthquake. J Geophys Res Solid Earth 125:1–32. https://doi.org/10.1029/2019JB018703

Hastings W (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109. https://doi.org/10.1093/biomet/57.1.97

Hoffman MD, Gelman A (2014) The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 15:1593–1623

Ito T, Gunawan E, Kimata F, Tabei T, Simons M, Meilano I, Agustan N, Ohta Y, Nurdin I, Sugiyanto D (2012) Isolating along-strike variations in the depth extent of shallow creep and fault locking on the northern Great Sumatran fault. J Geophys Res Solid Earth 117:B06409. https://doi.org/10.1029/2011JB008940

Kawamoto S, Hiyama Y, Ohta Y, Nishimura T (2016) First result from the GEONET real-time analysis system (REGARD): the case of the 2016 Kumamoto earthquakes. Earth Planets Space 68:190. https://doi.org/10.1186/s40623-016-0564-4

Kawamoto S, Ohta Y, Hiyama Y, Todoriki M, Nishimura T, Furuya T, Sato Y, Yahagi T, Miyagawa K (2017) REGARD: a new GNSS-based real-time finite fault modeling system for GEONET. J Geophys Res Solid Earth 122:1324–1349. https://doi.org/10.1002/2016JB013485

Link WA, Eaton MJ (2012) On thinning of chains in MCMC. Methods Ecol Evol 3:112–115. https://doi.org/10.1111/j.2041-210X.2011.00131.x

Mai PM, Beroza GC (2002) A spatial random field model to characterize complexity in earthquake slip. J Geophys Res Solid Earth 107:ESE10-1-ESE10-2. https://doi.org/10.1029/2001JB000588

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092. https://doi.org/10.1063/1.1699114

Minson SE, Murray JR, Langbein JO, Gomberg JS (2014) Real-time inversions for finite fault slip models and rupture geometry based on high-rate GPS data. J Geophys Res Solid Earth 119:3201–3231. https://doi.org/10.1002/2013JB010622

Muir JB, Tkalčić H (2020) Probabilistic lowermost mantle P-wave tomography from hierarchical Hamiltonian Monte Carlo and model parametrization cross-validation. Geophys J Int 223:1630–1643. https://doi.org/10.1093/gji/ggaa397

Musa A, Watanabe O, Matsuoka H, Hokari H, Inoue T, Murashima Y, Ohta Y, Hino R, Koshimura S, Kobayashi H (2018) Real-time tsunami inundation forecast system for tsunami disaster prevention and mitigation. J Supercomput 74:3093–3113. https://doi.org/10.1007/s11227-018-2363-0

Neal RM (2011) MCMC using Hamiltonian dynamics. Handbook of Markov chain Monte Carlo. Chapman & Hall/CRC, London

Nesterov Y (2009) Primal-dual subgradient methods for convex problems. Math Program 120:221–259. https://doi.org/10.1007/s10107-007-0149-x

Ohno K, Ohta Y, Kawamoto S, Abe S, Hino R, Koshimura S, Musa A, Kobayashi H (2021) Real-time automatic uncertainty estimation of coseismic single rectangular fault model using GNSS data. Earth Planets Space 73:127. https://doi.org/10.1186/s40623-021-01425-0

Ohno K, Ohta Y, Hino R, Koshimura S, Musa A, Abe T, Kobayashi H (2022) Rapid and quantitative uncertainty estimation of coseismic slip distribution for large interplate earthquakes using real-time GNSS data and its application to tsunami inundation prediction. Earth Planets Space 74:24. https://doi.org/10.21203/rs.3.rs-923956/v1

Ohta Y, Ohzono M, Miura S, Iinuma T, Tachibana K, Takatsuka K, Miyao K, Sato T, Umino N (2008) Coseismic fault model of the 2008 Iwate-Miyagi Nairiku earthquake deduced by a dense GPS network. Earth Planets Space 60:1197–1201. https://doi.org/10.1186/BF03352878

Ohta Y, Kobayashi T, Tsushima H, Miura S, Hino R, Takasu T, Fujimoto H, Iinuma T, Tachibana K, Demachi T, Sato T, Ohzono M, Umino N (2012) Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: application to the 2011 Tohoku-Oki earthquake (Mw 9.0). J Geophys Res Solid Earth 117:1–16. https://doi.org/10.1029/2011JB008750

Ohta Y, Inoue T, Koshimura S, Kawamoto S, Hino R (2018) Role of real-time GNSS in near-field tsunami forecasting. J Disaster Res 13:453–459. https://doi.org/10.20965/jdr.2018.p0453

Okada Y (1992) Internal deformation due to shear and tensile faults in a half-space. Bull Seismol Soc Am 82:1018–1040

Ringer H, Whitehead JP, Krometis J, Harris RA, Glatt-Holtz N, Giddens S, Ashcraft C, Carver G, Robertson A, Harward M, Fullwood J, Lightheart K, Hilton R, Avery A, Kesler C, Morrise M, Klein MH (2021) Methodological reconstruction of historical seismic events from anecdotal accounts of destructive tsunamis: a case study for the great 1852 Banda arc mega-thrust earthquake and tsunami. J Geophys Res Solid Earth 126:1–23. https://doi.org/10.1029/2020JB021107

Yamada *et al. Earth, Planets and Space*   (2022) 74:86

Page 16 of 16

Roberts GO, Rosenthal JS (1998) Optimal scaling of discrete approximations to Langevin diffusions. J R Stat Soc Ser B Methodol 60:255–268. https://doi.org/10.1111/1467-9868.00123

Sambridge M (2013) A parallel tempering algorithm for probabilistic sampling and multimodal optimization. Geophys J Int 196:357–374. https://doi.org/10.1093/gji/ggt342

Sambridge M, Mosegaard K (2002) Monte Carlo methods in geophysical inverse problems. Rev Geophys 40:3-1-3–29. https://doi.org/10.1029/2000RG000089

Swendsen RH, Wang JS (1986) Replica Monte Carlo simulation of spin-glasses. Phys Rev Lett 57:2607–2609. https://doi.org/10.1103/PhysRevLett.57.2607

Tanaka Y, Ohta Y, Miyazaki S (2019) Real-time coseismic slip estimation via the GNSS carrier phase to fault slip approach: a case study of the 2016 Kumamoto earthquake. Geophys Res Lett 46:1367–1374. https://doi.org/10.1029/2018GL080741

Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, Philadelphia

Yarai H, Kobayashi T, Morishita Y, Fujiwara S, Hiyama Y, Kawamoto S, Ueshiba H, Miura Y, Miyahara B (2016) Source fault models of the 2016 Kumamoto earthquake inverted from crustal deformation. Bull Geospat Inf Auth Jpn 128:169–176 (**in Japanese**)

Publisher's Note