AGU100 ADVANCING EARTH AND SPACE SCIENCE

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE

10.1002/2017JB015249

Key Points:
- We present Bayesian inversion for point-localized wavefield sources
- In numerical examples, HMC only requires around 100 samples for good approximation of the posterior
- Spectral element simulations are used to account for 3-D structure from full-waveform inversion

Correspondence to:
A. Fichtner,
andreas.fichtner@erdw.ethz.ch

Citation:
Fichtner, A., & Simutė, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. *Journal of Geophysical Research: Solid Earth*, *123*. https://doi.org/10.1002/2017JB015249

Received 17 NOV 2017
Accepted 13 MAR 2018
Accepted article online 23 MAR 2018

©2018. The Authors.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Hamiltonian Monte Carlo Inversion of Seismic Sources in Complex Media

Andreas Fichtner[1] **and Saule Simutė**[1]

[1] Institute of Geophysics, ETH Zurich, Zurich, Switzerland

Abstract We present a probabilistic seismic point source inversion, taking into account 3-D heterogeneous Earth structure. Our method rests on (1) reciprocity and numerical wavefield simulations in complex media and (2) Hamiltonian Monte Carlo sampling that requires only a small amount of test models to provide reliable uncertainty information on the timing, location, and mechanism of the source. Using spectral element simulations of 3-D, viscoelastic, anisotropic wave propagation, we precompute receiver side strain tensors in time and space. This enables the fast computation of synthetic seismograms for any hypothetical source within the volume of interest, and thus a Bayesian solution of the inverse problem. To improve efficiency, we developed a variant of Hamiltonian Monte Carlo sampling. Taking advantage of easily computable derivatives, numerical examples indicate that Hamiltonian Monte Carlo can converge to the posterior probability density with orders of magnitude less samples than the derivative-free Metropolis-Hastings algorithm, which we use for benchmarking. Exact numbers depend on observational errors and the quality of the prior. We apply our method to the Japanese Islands region where we previously constrained 3-D structure of the crust and upper mantle using full-waveform inversion with a minimum period of 15 s.

1. Introduction

The characterization of effectively point-localized wavefield sources has been a prime objective of seismology ever since the first installations of sufficiently dense seismometer networks. Knowing the location, timing, moment tensor, and possibly the source time function of earthquakes is critical for a wide range of related fields and applications. These include seismotectonics, earthquake physics, seismic hazard analysis, early warning, nuclear monitoring, and seismic tomography.

1.1. Developments and Challenges in Seismic Source Inversion

Pioneering work on the mathematical description of earthquake sources (e.g., Backus & Mulcahy, 1976a, 1976b; Burridge & Knopoff, 1964; Knopoff & Randall, 1970) was accompanied by the development of practical methods to constrain their properties (e.g., Aki & Patton, 1978; Buland & Gilbert, 1976; Dziewoński & Gilbert, 1974; Dziewoński & Woodhouse, 1981; Dziewoński et al., 1981; Gilbert, 1973; Randall & Knopoff, 1970). As a result, routinely delivered source locations and moment tensors have become an indispensable and routinely employed element of geophysical research.

Despite the conceptual simplicity of the seismic source inverse problem, methods for the characterization of point-localized sources continue to proliferate, focusing on near-real-time operation and automation (e.g., Bernardi et al., 2004; Cesca et al., 2010; Käufl et al., 2015; Scognamiglio et al., 2009; Vackár et al., 2017), uncertainty quantification (e.g., Mustać & Tkalčić, 2016; Silwal & Tape, 2016; Staehler & Sigloch, 2014, 2017; Valentine & Trampert, 2012; Wéber, 2006), the incorporation of 3-D structural models (e.g., Chen et al., 2007; Fichtner & Tkalcic, 2010; Hejrani et al., 2017; Hingee et al., 2011; Hsieh et al., 2014; Liu et al., 2004; Zhu & Zhou, 2016), and the use of new types of seismological observables (e.g., Donner et al., 2016; O'Toole et al., 2012).

These continued developments also reflect fundamental difficulties in seismic source inversion. They include trade-offs between source parameters and 3-D Earth structure, and the existence of a null space, which often contains, for instance, the trace of the moment tensor (e.g., Dufumier & Rivera, 1997). The unavoidable null space commands the use of Monte Carlo methods to sample the space of acceptable solutions and to avoid results that are biased by regularization (e.g., Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002).

Unfortunately, Monte Carlo methods for seismic source inversion suffer particularly strongly from the curse of dimensionality (e.g., Tarantola, 2005) for two reasons: (1) Since source parameters, such as moment tensor components, may vary over several orders of magnitude, the prior in model space tends to be very wide. In other words, the space (or number) of models that are a priori plausible is very large. (2) Since the observational errors in today's seismic data are often very small, the posterior in model space is rather concentrated, meaning that few models are actually plausible when tested against high-quality data.

As a consequence, many samples that are tested for being a priori plausible may in fact be wasted because they fail to produce synthetic seismograms that explain observed seismograms to within the small measurement uncertainties. This facet of the curse of dimensionality severely reduces the efficiency of Monte Carlo methods that strongly rely on prior information, including the widely used Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953; Tarantola, 2005).

1.2. Hamiltonian Monte Carlo

Motivated by the potentially poor performance of Monte Carlo methods for high-dimensional problems, Hamiltonian Monte Carlo was introduced as hybrid Monte Carlo by Duane et al. (1987) in the context of lattice quantum chromodynamics. The method moved into the focus of statistical computation through the work of Neal (1996) who used it for Bayesian neural network learning. Hamiltonian Monte Carlo was further popularized by the reviews of Neal (2011) and Betancourt (2017), who introduced the method without the overhead of advanced differential geometry. Today, applications of Hamiltonian Monte Carlo can be found in a wide range of disciplines, including neural networks and machine learning (e.g., Bishop, 2006), molecular simulations (e.g., Dubbledam et al., 2016), nuclear physics (e.g., Elhatisari et al., 2015), genomics (e.g., Honkela et al., 2015), and quantum mechanics (e.g., Seah et al., 2015). However, despite its success in widely varying disciplines, Hamiltonian Monte Carlo seems not to be used for the solution of geophysical inverse problems, with few recent exceptions (e.g., Muir & Tkalčić, 2015; Sen & Biswas, 2017).

1.3. Objectives and Outline

This work has two main objectives: (1) to introduce Hamiltonian Monte Carlo to geophysical inversion using the low-dimensional point source inversion problem as example and (2) to illustrate the efficiency of Hamiltonian Monte Carlo for seismic source inversion, taking into account 3-D heterogeneous Earth structure. Less explicitly, we further intend to pave the way toward higher-dimensional finite-source inversions where the benefits of Hamiltonian Monte Carlo are expected to become more pronounced than in the relatively low-dimensional point source problem.

This manuscript is organized as follows: In section 2 we summarize the forward problem solution, which rests on precomputed, receiver side Green's functions. Section 3 introduces Hamiltonian Monte Carlo in the context of Bayesian inference, with special focus on weakly nonlinear problems. A seismic source inversion toy problem that illustrates the basic concepts is the subject of section 4. Finally, in section 5, we present a real-data point source inversion from the Japanese Islands region.

2. Forward Problem Solution

To set the stage for the probabilistic inversion, we describe the forward problem solution. For a moment tensor point source at position $\boldsymbol{\xi}$, the i component of the displacement field at position $\mathbf{x}$ and time t is given by the representation theorem (e.g., Aki & Richards, 2002) as

$$u_i(\mathbf{x}, t) = \sum_{n,j=1}^{3} \int M_{nj}(t - t') \frac{\partial}{\partial \xi_j} G_{in}(\mathbf{x}, t'; \boldsymbol{\xi})\, \mathrm{d}t' \,. \tag{1}$$

In equation (1), M_{nj} denotes the components of the symmetric moment tensor, and $G_{in}(\mathbf{x}, t'; \boldsymbol{\xi})$ is the i component of the Green's function for a point source in n direction at position $\boldsymbol{\xi}$. Equation (1) is inconvenient because the simulation of G_{in} for heterogeneous media from all potential sources $\boldsymbol{\xi}$ is prohibitively expensive. We therefore employ spatial reciprocity (e.g., Aki & Richards, 2002) to modify the representation of u_i to

$$u_i(\mathbf{x}, t) = \sum_{n,j=1}^{3} \int M_{nj}(t - t') \frac{\partial}{\partial \xi_j} G_{ni}(\boldsymbol{\xi}, t'; \mathbf{x})\, \mathrm{d}t' \,. \tag{2}$$

This allows us to model u_i via the precomputation of a database containing the derivatives of the receiver side Green's functions, $\partial G_{ni}(\xi, t'; \mathbf{x})/\partial \xi_j$, for all receiver locations $\mathbf{x}$. Equation (2) has been used in various source studies (e.g., Chen et al., 2007; Hejrani et al., 2017; Lee et al., 2014; Zhao et al., 2006).

To enable source inversion, we parametrize the moment tensor in terms of time-independent tensor components M_{nj}, the origin time t_0, and a small number of coefficients s_k that premultiply the basis functions $\phi_k(t)$ of the source time function (moment function),

$$M_{nj}(t) = M_{nj} \sum_{k=1}^{N} s_k \phi_k(t - t_0) \,. \tag{3}$$

In the interest of generality, we leave the shape of the basis functions $\phi_k(t)$ unspecified at this point. Based on equation (3), the forward problem solution depends on $10 + N$ parameters, six independent components of the moment tensor M_{nj}, three source coordinates $\xi = (\xi_1, \xi_2, \xi_3)$, the origin time t_0, and an adjustable number N of source time function coefficients s_k. While alternative parametrizations of the time-independent tensor components have been proposed (e.g., Jost & Herrmann, 1989; Kikuchi & Kanamori, 1991; Staehler & Sigloch, 2014; Tape & Tape, 2015), we prefer to work directly with M_{nj} in the interest of simplicity and because it preserves the linear dependence of synthetic seismograms on the source mechanism. This aspect will be further discussed in section 6.2.

Equations (2) and (3) define the forward problem, that is, the computation of synthetic seismograms as a function of model parameters. In the following paragraphs we describe how these model parameters and their uncertainties can be inferred efficiently using Hamiltonian Monte Carlo inversion.

3. Hamiltonian Monte Carlo

3.1. General Hamiltonian Monte Carlo Sampling

For notational convenience, we collect all model parameters that we wish to infer into an N_q-dimensional vector $\mathbf{q}$. Thus, $\mathbf{q}$ comprises moment tensor components, source coordinates, origin time, and source time function coefficients. Our goal is to sample the posterior probability density $\pi(\mathbf{q}|\mathbf{d})$ given by Bayes' theorem as

$$\pi(\mathbf{q}|\mathbf{d}) = k\, \pi(\mathbf{d}|\mathbf{q})\, \pi(\mathbf{q}) \,, \tag{4}$$

where $\mathbf{d}$ is the data vector, $\pi(\mathbf{d}|\mathbf{q})$ is the likelihood function, $\pi(\mathbf{q})$ is the model space prior, and k is a normalization constant (e.g., Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002; Tarantola, 2005). In the context of Hamiltonian Monte Carlo, a model $\mathbf{q}$ is interpreted as the position vector of a particle, and equation (4) is used to define its potential energy U as the negative natural logarithm of the posterior,

$$U(\mathbf{q}) = -\ln\, \pi(\mathbf{q}|\mathbf{d}) \,. \tag{5}$$

Thus, more plausible models $\mathbf{q}$ with large values of the posterior correspond to low potential energies, and vice versa. The kinetic energy K of the model $\mathbf{q}$ is defined as an auxiliary variable through the artificially introduced N_q-dimensional momentum vector $\mathbf{p}$,

$$K(\mathbf{p}) = \frac{1}{2} \mathbf{p}^T \mathcal{M}^{-1} \mathbf{p} \,. \tag{6}$$

The $N_q \times N_q$ mass matrix $\mathcal{M}$ is a tuning parameter. While it can in principle be set to any arbitrary positive definite matrix, its choice affects the performance of the Hamiltonian Monte Carlo algorithm, as we will explain in section 3.2. Endowed with momentum, the model or particle $\mathbf{q}$ will move through the $2N_q$-dimensional phase space according to Hamilton's equations (e.g., Landau & Lifshitz, 1976; Symon, 1971),

$$\frac{\mathrm{d}q_i}{\mathrm{d}\tau} = \frac{\partial K}{\partial p_i} \,, \qquad \frac{\mathrm{d}p_i}{\mathrm{d}\tau} = -\frac{\partial U}{\partial q_i} \,, \quad i = 1, \ldots, N_q \,, \tag{7}$$

and with increasing artificial time τ that is not to be confused with the physical time t. Based on this notion of a model moving as a particle through phase space, the Hamiltonian Monte Carlo algorithm operates in the following sequence of steps, starting from a randomly chosen initial model $\mathbf{q}$:

1. Randomly draw the N_q momenta p_i from the normal distribution $\exp\left(-\frac{1}{2}\mathbf{p}^T \mathcal{M}^{-1} \mathbf{p}\right)$.
2. With position and momentum thus determined, propagate the model $\mathbf{q}$ forward in time by solving Hamilton's equations (7). After some time τ, which remains to be chosen, a new model $\mathbf{q}(\tau)$ with a new momentum $\mathbf{p}(\tau)$ is reached.

3. The new model is accepted with probability

$$\Pi_{\text{accept}} = \min\left[1, \frac{\exp\left[-H(\mathbf{p}(\tau), \mathbf{q}(\tau))\right]}{\exp\left[-H(\mathbf{p}, \mathbf{q})\right]}\right], \tag{8}$$

where $H = K + U$ is the total energy or Hamiltonian of the model. If accepted, $\mathbf{q}(\tau)$ will serve as initial model for the next trajectory. Otherwise, $\mathbf{q}$ will be reused as starting point.

4. Return to (1) and repeat until sufficiently many samples have been produced.

The models generated by the algorithm are samples of the posterior $\pi(\mathbf{q}|\mathbf{d})$ and may be used to compute marginals, means, covariances, or other quantities of interest. Tuning parameters that affect convergence are the mass matrix $\mathcal{M}$ and the propagation time τ along the Hamiltonian trajectory. These will be discussed below in the context of specific applications.

3.2. Weakly Nonlinear Problems

Most of the computational expense in Hamiltonian Monte Carlo is in the evaluation of $\partial U/\partial q_i$ along the Hamiltonian trajectory, because this requires us to compute derivatives of the posterior and, therefore, of the forward problem. The computation cost can be reduced substantially in the case of Gaussian observational errors and weakly nonlinear problems where the forward equations can be meaningfully linearized. In fact, under the Gaussian assumption and in the context of seismic source inversion, the potential energy U defined in equation (5) is equal to

$$U(\mathbf{q}) = \underbrace{\frac{1}{2T}\sum_{r=1}^{N_r}\sum_{n=1}^{3}\int_0^T \sigma_d^{-2}\left[u_n(\mathbf{x}_r, t; \mathbf{q}) - u_n^{\text{obs}}(\mathbf{x}_r, t)\right]^2 dt}_{U_d(\mathbf{q})} + \underbrace{\frac{1}{2N_q}(\mathbf{q}-\mathbf{q}^0)^T\mathbf{C}_q^{-1}(\mathbf{q}-\mathbf{q}^0)}_{U_q(\mathbf{q})}. \tag{9}$$

The first term in equation (9), U_d, is the L_2 waveform misfit between observations $u_n^{\text{obs}}(\mathbf{x}_r, t)$ and synthetics $u_n(\mathbf{x}_r, t; \mathbf{q})$ at N_r receiver positions $\mathbf{x}_r$, integrated over the time interval $[0, T]$. The data covariance σ_d^2 encodes the observational uncertainties. While σ_d^2 could be made time dependent and receiver dependent, we keep it constant for notational convenience. The second term in equation (9), U_q, captures Gaussian prior knowledge on the model parameters $\mathbf{q}$, with prior mean $\mathbf{q}^0$ and prior covariance matrix $\mathbf{C}_q$. Using equations (5) and (9), the likelihood function and the model-space prior may be written as

$$\pi(\mathbf{d}|\mathbf{q}) = \exp\left[-U_d(\mathbf{q})\right], \quad \pi(\mathbf{q}) = \exp\left[-U_q(\mathbf{q})\right]. \tag{10}$$

Expanding the synthetics in equation (9) around $\mathbf{q}^0$ gives correct to first order,

$$u_n(\mathbf{x}_r, t; \mathbf{q}) = u_n(\mathbf{x}_r, t; \mathbf{q}^0) + \sum_{i=1}^{N_q}\frac{\partial}{\partial q_i}u_n(\mathbf{x}_r, t; \mathbf{q}^0)\,(q_i - q_i^0). \tag{11}$$

With the forward problem defined in equations (2) and (3), expression (11) is exact for the moment tensor components M_{nj} and the source time function coefficients s_k. Higher-order terms only exist for the source coordinates ξ and the origin time t_0. Substituting (11) into (9) and isolating the model parameters q_i yields a simplified expression for the potential energy,

$$U(\mathbf{q}) = \frac{1}{2}\sum_{i,j=1}^{N_q}(q_i - q_i^0)\,A_{ij}\,(q_j - q_j^0) + \sum_{i=1}^{N_q} b_i(q_i - q_i^0) + \frac{1}{2}c. \tag{12}$$

The quantities A_{ij}, b_i, and c are defined as

$$\begin{aligned}
A_{ij} &= \frac{1}{T\sigma_d^2}\sum_{r=1}^{N_r}\sum_{n=1}^{3}\int_0^T \frac{\partial}{\partial q_i}u_n(\mathbf{x}_r, t; \mathbf{q}^0)\,\frac{\partial}{\partial q_j}u_n(\mathbf{x}_r, t; \mathbf{q}^0)\,\mathrm{d}t + \frac{1}{N_q\sigma_q^2},\\
b_i &= \frac{1}{T\sigma_d^2}\sum_{r=1}^{N_r}\sum_{n=1}^{3}\int_0^T \left[u_n(\mathbf{x}_r, t; \mathbf{q}^0) - u_n^{\text{obs}}(\mathbf{x}_r, t)\right]\,\frac{\partial}{\partial q_i}u_n(\mathbf{x}_r, t; \mathbf{q}^0)\,\mathrm{d}t,\\
c &= \frac{1}{T\sigma_d^2}\sum_{r=1}^{N_r}\sum_{n=1}^{3}\int_0^T \left[u_n(\mathbf{x}_r, t; \mathbf{q}^0) - u_n^{\text{obs}}(\mathbf{x}_r, t)\right]^2\,\mathrm{d}t.
\end{aligned} \tag{13}$$

Being independent of the variable $\mathbf{q}$, they can be precomputed before running the Hamiltonian Monte Carlo sampling. With the help of (12), the derivative of the potential energy is reduced to the simple vector-matrix product,

$$\frac{\partial U}{\partial q_i} = \sum_{j=1}^{N_q} A_{ij}(q_j - q_j^0) + b_i \,, \tag{14}$$

which does not require significant computational resources. The first-order approximation (11) leads to a modified Hamiltonian Monte Carlo algorithm where an approximate trajectory is computed on the basis of equation (14), while the correct criterion (8) is still used to decide about the acceptance of new samples.

In addition to yielding a computationally more efficient algorithm, equation (14) can also guide the choice of the mass matrix $\mathcal{M}$ that we have so far left unspecified. Indeed, substituting (14) into Hamilton's equations (7) gives a second-order differential equation for $\mathbf{q}(\tau)$,

$$\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}\tau^2} + \mathcal{M}^{-1}\mathbf{A}(\mathbf{q} - \mathbf{q}^0) = -\mathcal{M}^{-1}\mathbf{b} \,. \tag{15}$$

Using the matrix exponential and matrix square root, the solution of (15) can be symbolically written as

$$\mathbf{q}(\tau) = \exp\left(i\sqrt{\mathcal{M}^{-1}\mathbf{A}\tau}\right)\mathbf{q}_1 + \mathbf{q}_2 \,, \tag{16}$$

with some constant vectors $\mathbf{q}_1$ and $\mathbf{q}_2$. Equation (16) reveals that the model $\mathbf{q}(\tau)$ oscillates through phase space along closed Hamiltonian trajectories. Depending on the product $\mathcal{M}^{-1}\mathbf{A}$, some components of $\mathbf{q}(\tau)$ may oscillate very rapidly and explore phase space in a short time, while others may oscillate slowly and take a long time for phase space oscillation. However, for an efficient algorithm it is desirable that all components oscillate with roughly equal speed. This can be achieved by choosing $\mathcal{M} = \mathbf{A}$ when $\mathbf{A}$ is indeed a positive definite covariance matrix that can be used to draw random momenta in Step 1 of the Hamiltonian Monte Carlo algorithm. Otherwise, choosing $\mathcal{M}$ to be a diagonal matrix with entries $Q_{ii} = M_{ii}$ empirically works well in the examples presented below.

In the following sections we will illustrate the application of the modified Hamiltonian Monte Carlo algorithm, first using an easily reproducible toy problem, and then with a real-data application in the Japanese Islands region.

3.3. Numerical Details

To integrate Hamilton's equations (7), we use the leapfrog algorithm. Being symplectic, it ensures that the time-discrete equations still maintain two properties of Hamiltonian dynamics that are essential for Hamiltonian Monte Carlo, volume preservation and reversability (Neal, 2011). Following Neal (1996), we choose the integration step size empirically to be slightly less than the minimum step size required for numerical stability. This ensures that the integration error remains moderate and that the phase space is explored efficiently with the minimum number of integration steps. The length of individual Hamiltonian trajectories, that is, τ in equation (8), is found adaptively using the No-U-Turn criterion (Hoffmann & Gelman, 2014). This ensures that trajectories terminate when they begin to return toward their starting point. While a mathematical proof for the efficiency of the No-U-Turn criterion still seems to be missing, it is commonly found to work well empirically (Betancourt, 2017).

4. Toy Problem Illustrations

We begin our illustrative tour with simplistic synthetic inversions of *P* waveforms in a homogeneous elastic full space where the forward problem can be solved analytically (e.g., Aki & Richards, 2002). These examples are intended to be educational and to explain key features of Hamiltonian Monte Carlo.

4.1. Model Space Exploration

In the first example we only consider the moment tensor components M_{ij}, keeping all other parameters fixed. The source is at $(x_s, y_s, z_s) = (0, 0, 0)$ m with $M_{xx} = 1$ and $M_{yy} = M_{zz} = M_{xy} = M_{xz} = M_{yz} = 0$ Nm. Three receivers are located along the coordinate axes at $(x_1, y_1, z_1) = (10^3, 0, 0)$, $(x_2, y_2, z_2) = (0, 10^3, 0)$, and $(x_3, y_3, z_3) = (0, 0, 10^3)$ m. In this configuration, only the diagonal components M_{xx}, M_{yy}, and M_{zz} can be constrained. The off-diagonals M_{xy}, M_{xz}, and M_{yz} are purely controlled by the model-space prior $\pi(\mathbf{q})$ from equations (9) and (10). For simplicity, we use a zero prior mean $\mathbf{q}^0 = \mathbf{0}$ and a diagonal prior covariance matrix $\mathbf{C}_d = \sigma_d^2\mathbf{I}$,

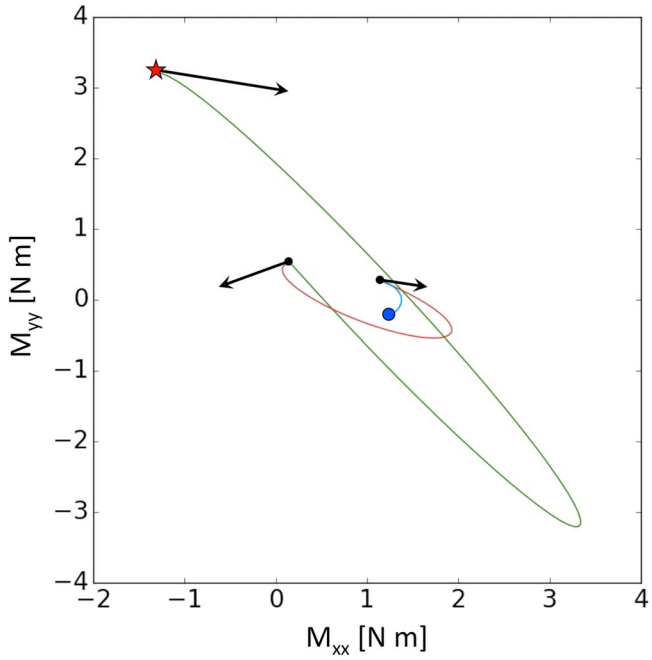

Figure 1. Two-dimensional projection onto the M_{xx}-M_{yy} plane of the first three trajectories of a Hamiltonian Monte Carlo run. The red star and the blue dot mark the start and end points, respectively. Black arrows symbolize the random momentum vectors at the beginning of the trajectories, plotted in different colors.

where **I** is the identity matrix. Both the prior standard deviation, σ_q, and the observational errors σ_d we will vary during the subsequent numerical experiments.

To understand the functioning of Hamiltonian Monte Carlo, we consider the projection onto the M_{xx}-M_{yy} plane of the first three trajectories of a representative random walk, shown in Figure 1. The walk starts at a random position, indicated by the red star. There, the model **q** is given a random momentum **p**, shown by a black arrow. Following Hamiltonian dynamics, the model moves along an elliptical trajectory, roughly in the direction of $M_{xx} = 1$ and $M_{yy} = 0$ Nm. At the end of the trajectory, the model receives a new random momentum, which takes it onto a different elliptical trajectory, roughly orbiting around $M_{xx} = 1$ and $M_{yy} = 0$ Nm. The same procedure is then repeated. A key feature of Hamiltonian Monte Carlo, seen from this illustration, is that models stay within the *typical set*, that is, the relevant part of the model space where the posterior $\pi(\mathbf{q}|\mathbf{d})$ is large. This property is essential for the efficiency of Hamiltonian Monte Carlo. It ensures that new proposals are likely to be accepted and that the model space is explored rapidly.

The importance of a high acceptance rate is illustrated in Figure 2, where we compare Hamiltonian Monte Carlo to the Metropolis-Hastings algorithm. The latter proposes samples randomly according to the prior $\pi(\mathbf{q})$ (Hastings, 1970; Metropolis et al., 1953; Mosegaard & Tarantola, 1995). The prior standard deviation is set to $\sigma_q = 0.5$ Nm for all moment tensor components, and the observational errors σ_d decrease successively. (During each individual inversion, σ_d is constant, but variable from one inversion to the next.) When σ_d is as large as the *P* wave amplitude itself ($\sigma_d = 1$ in Figure 2a), the posterior $\pi(\mathbf{q}|\mathbf{d})$ is broad because many models fit the artificial observations acceptably well. Therefore, more than 60% of the test samples proposed during the Metropolis-Hastings algorithm are accepted, and the model space is explored efficiently. However, as the observational error σ_d decreases, the percentage of accepted test models drops superexponentially. When σ_d equals 5% of the maximum *P* wave amplitude, not uncommon in practice, only 0.15% of all proposals are accepted. The algorithm is stalled at few models because the probability of randomly drawing a test model within the very small typical set of the posterior has become negligible. Consequently, the model space is explored very inefficiently, and most of the computational resources are wasted. In contrast to Metropolis-Hastings, the acceptance rate of Hamiltonian Monte Carlo is constant around 40%, independent of the observational errors.

The drop in acceptance rate observed in Figure 2a is a consequence of shrinking the typical set of the posterior relative to the fixed typical set of the prior, as observational uncertainties decrease. A similar phenomenon occurs when the typical set of the prior expands relative to a fixed typical set of the posterior. This is shown in Figure 2b where the standard deviation of the prior σ_q increases from 0.5 to 2.5 Nm. (Again, σ_q is constant within each inversion but varies from one run to the next.) The increasingly weaker prior enlarges the typical set of the prior from which the Metropolis-Hastings algorithm draws its proposals. Consequently, it becomes less likely to obtain a proposal from the typical set of the posterior, and the acceptance rate drops. Again, Hamiltonian Monte Carlo does not suffer from this problem, which becomes even more severe in higher dimensions as a consequence of the *curse of dimensionality* (e.g., Tarantola, 2005).

In summary, Figure 2 illustrates that Hamiltonian Monte Carlo is well suited for scenarios that are typical for seismic source inversion: high data quality (low observational errors) and weak prior knowledge on parameters such as moment tensor components that may vary over several orders of magnitude.

4.2. Convergence

The efficient exploration of model space in Hamiltonian Monte Carlo favorably affects convergence, as illustrated in Figure 3 for the case of $\sigma_d = 0.1$ and $\sigma_q = 2.0$ Nm. Using 20 bins, the posterior marginal of M_{xx} requires several thousand samples to converge. However, the posterior mean and the posterior standard deviation practically reach their final values already after a short burn-in phase of several tens of samples, needed to find the typical set.

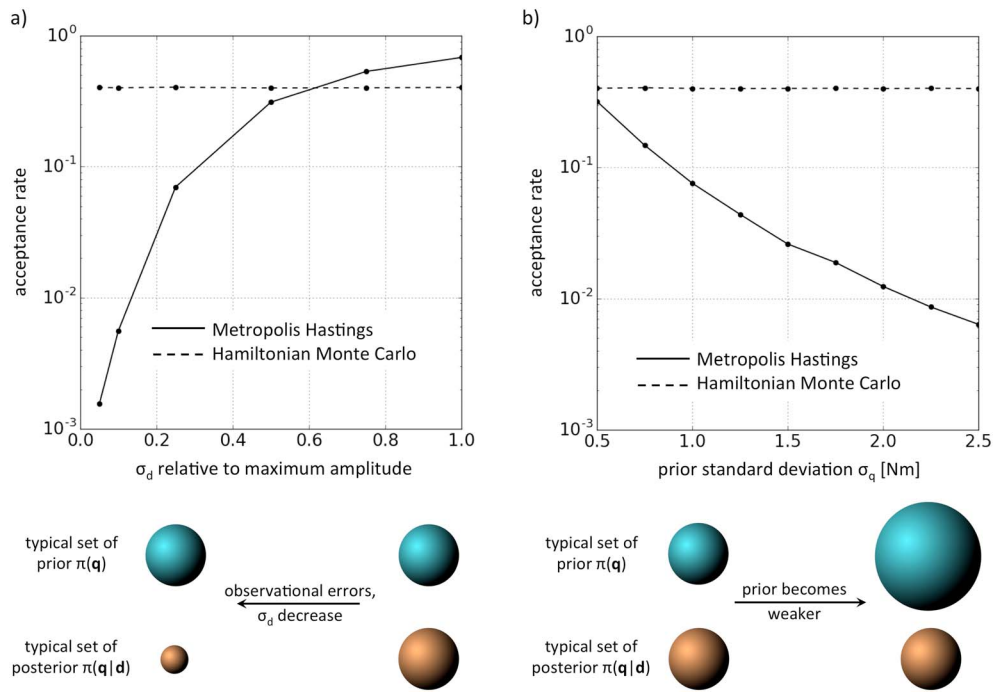

Figure 2. Acceptance rate of the Metropolis-Hastings algorithm (solid) and Hamiltonian Monte Carlo (dashed) averaged over 10 runs with 10,000 samples each. (a) Acceptance rate as a function of the observational error σ_d in equation (9), normalized to the maximum amplitude of the *P* wave. For Metropolis-Hastings, the acceptance rate drops superexponentially with decreasing observational errors, while the acceptance rate for Hamiltonian Monte Carlo is constant at 40%. (b) Acceptance rate as a function of the prior standard deviation σ_q for the moment tensor components. For an increasingly weaker prior, that is, increasing prior standard deviation, the acceptance rate for Metropolis-Hastings decreases rapidly, while the acceptance rate of Hamiltonian Monte Carlo is again unaffected. The behavior of the prior and posterior typical sets for decreasing errors and weaker priors is illustrated in color below.

For the same choice of σ_d and σ_q, the acceptance rate of the Metropolis-Hastings algorithm is on the order of 10^{-4}, meaning that any consideration of convergence would not even be meaningful when the number of samples is not some orders of magnitude larger than 10^4. While Figure 3 only shows the marginal of M_{xx}, we note that maginals for the remaining moment tensor components behave identically.

4.3. Extension to Source Location, Origin Time, and Source Time Function

Based on the formulation developed in section 3.2, the previous example can be extended to include origin time, source location, and source time function coefficients as free parameters. As bases for the source time function, ϕ_k, we choose 10 nonoverlapping box functions of duration 1 s. The dimension of the model space is therefore 20.

Convergence and acceptance rate of Hamiltonian Monte Carlo behave similar as in the reduced scenario shown in Figures 2 and 3. This suggests that the linear approximation for origin time and source location indeed works well. For illustration, we show a collection of 2-D posterior marginals for different parameter pairs, as well as 1-D posterior marginals for the source time function coefficients, in Figure 4. Notable, and intuitively plausible, features of the marginals for this simple example, such as the trade-off between source location x_s and origin time t_0, indicate that the algorithm functions as expected and is ready for real-data applications.

5. Application

To test the Hamiltonian Monte Carlo source inversion on observed waveforms, we consider an event from near the Izu arc, south of Japan (Figure 5a). In the following paragraphs, we will describe the data, the structural model used to compute synthetic seismograms, and the inferred properties of the event.

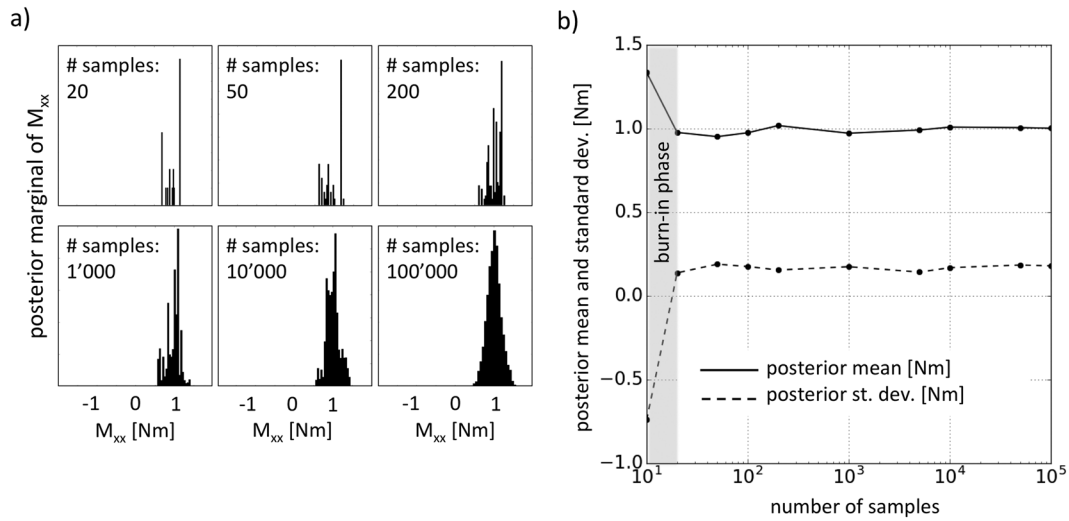

Figure 3. Convergence of the posterior marginal of the moment tensor component M_{xx} with increasing number of samples. (a) Histogram of the posterior marginal with 20 bins and a number of samples ranging between 20 and 100,000. (b) Convergence of the posterior mean and the posterior standard deviation of M_{xx}. While the posterior distribution in (a) requires several thousand samples to achieve acceptable convergence, the posterior mean and the posterior standard deviation converge after a very short burn-in phase of several tens of samples.

5.1. Data and Structural Model

The event used for this demonstration occurred on 30 May 2005. According to the global centroid moment tensor (CMT) catalog (www.globalcmt.org, last accessed on 17 November 2017), its origin was at latitude 32.20°, longitude 141.01°, depth 50.0 km, and GMT 11:18:08.47. The estimated magnitude was M_w 5.2. Among the available stations we selected 14 with an estimated signal-to-noise ratio above 10 in the period range of 15–100 s. The distribution of stations and the CMT location of the event are shown in Figure 5a.

The structural model is based on the full-waveform inversion of the Japanese Islands region by Simute et al. (2016). For the construction of the model, we used the numerical wave propagation code SES3D (Fichtner et al., 2009; Gokhberg & Fichtner, 2016), which combines spectral element simulations of wave propagation (Faccioli et al., 1997; Fichtner, 2010; Komatitsch & Vilotte, 1998) with adjoint techniques (e.g., Fichtner et al., 2006; Tarantola, 1988; Tromp et al., 2005). The waveform inversion framework LASIF (Krischer et al., 2015) was used for data management and for the iterative minimization of time- and frequency-dependent phase misfits between observed and synthetic seismograms (Fichtner et al., 2008). In preparation for this source inversion study, we further improved the model of Simute et al. (2016) by reducing the minimum period from 20 to 15 s, using the same full-waveform inversion method as before. The model is radially anisotropic and captures both crustal and mantle structure. To illustrate the complexity of the model, horizontal slices through the SV velocity distribution and 15 and 70 km depth are shown in Figure 5b.

5.2. Setup of the Hamiltonian Monte Carlo Sampling

To prepare for the posterior sampling, we chose the CMT solution as prior mean, $\mathbf{q}^0$, for the Taylor expansion in equation (11), and for the computation of the quantities A_{ij}, b_j, and c in equation (13). Since the duration of magnitude $\sim$ 5 earthquakes is usually only a few seconds (e.g., Vallée, 2013), which is small compared to the minimum period of 15 s, we restrict the model parameter space to the moment tensor components, the source location, and the origin time. Omitting the source time function, we therefore have $N_q = 10$. Figure 6 shows the prior source model, and a comparison of observed and prior synthetic seismograms computed with the structural model presented in the previous paragraph. While only the E-W components are shown, all three components are used in the analysis.

To avoid nonlinearity related to cycle skipping, we restrict measurements to time intervals where observed and prior synthetic seismograms match to within a half-cycle. This windowing may be relaxed when only inverting for the moment tensor components, which enter the forward problem purely linearly. We conservatively set the observational error to $\sigma_d = 0.1\mu$m. The prior model covariance $\mathbf{C}_q$ is diagonal with standard

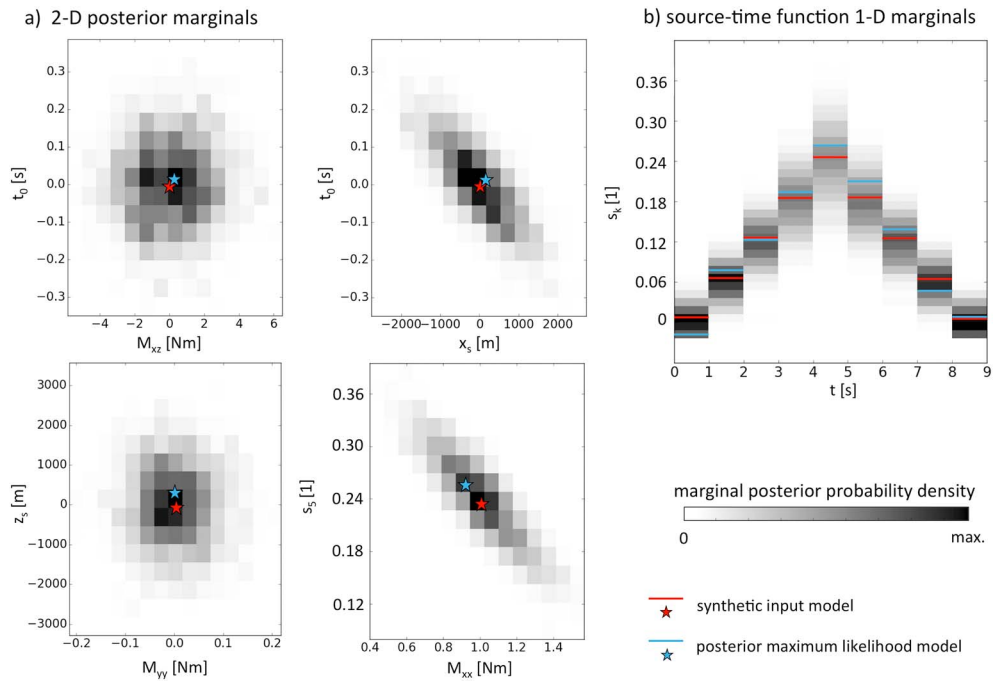

Figure 4. Selection of posterior marginals for the combined inference of source location, origin time, and source time function. (a) Trade-offs between pairs of model parameters are quantified by 2-D posterior marginals. The synthetic input model and the posterior maximum-likelihood model are indicated by red and blue stars, respectively. (b) Posterior marginal for the source time function coefficients. Prior and posterior maximum-likelihood values are marked by red and blue lines, respectively.

deviations set to infinity, thus reflecting a state of ignorance where the influence of the prior is negligible. Mathematically, $\sigma_q \to \infty$ eliminates the term $N_q^{-1}\sigma_q^{-2}$ in the computation of A_{ij} (equation (13)).

With the goal to assess convergence of the Hamiltonian Monte Carlo sampling, we perform two runs, the first with 1 million samples, the second with only 100 samples. The posterior maximum-likelihood model obtained from 100 samples is visualized in Figure 6a. The corresponding synthetic seismograms are shown below in Figure 6b by solid black curves. Despite the small number of samples, the maximum-likelihood source provides a significantly improved fit to the observations, especially at stations GJM, HID, HJO, HRO, and KMU. At stations HID and KMU the fit also improves visibly outside the actual measurement window.

5.3. Posterior Distributions

The quality of the maximum-likelihood model is fully characterized by the posterior, $\pi(\mathbf{q}|\mathbf{d})$, which constitutes the complete solution to the Bayesian inference or probabilistic inverse problem. Figure 7 visualizes the posterior in the form of 1-D marginals for the individual model parameters.

Standard deviations of the moment tensor components vary significantly, ranging from $\sim 0.4 \cdot 10^{16}$ Nm for M_{xy} to $\sim 1.5 \cdot 10^{16}$ Nm for M_{xx}. The uncertainties in the components of $\mathbf{M}$ translate into uncertainties of derived parameters, including the seismic moment and the trace, shown in Figure 7b. In fact, the maximum-likelihood moment tensor has a nonzero trace of $\sim 3 \cdot 10^{16}$ Nm, but a moment tensor with vanishing trace is only within little more than a standard deviation. Uncertainties of the time and space location parameters in Figure 7c are comparatively small, and the posterior maximum-likelihood is close to the prior mean. As a consequence of stations HJO and AOG being just slightly west of the event (Figure 5), longitude is better constrained than latitude.

A remarkable aspect is the number of samples needed to obtain useful approximations of the posterior mean and standard deviations for the individual parameters. These are indicated in Figure 7 for 1 million samples in blue and for 100 samples in red. While they differ, of course, by a few percent, they are for practical purposes identical. This important result confirms the preliminary synthetic study presented in section 4.2 and Figure 3.

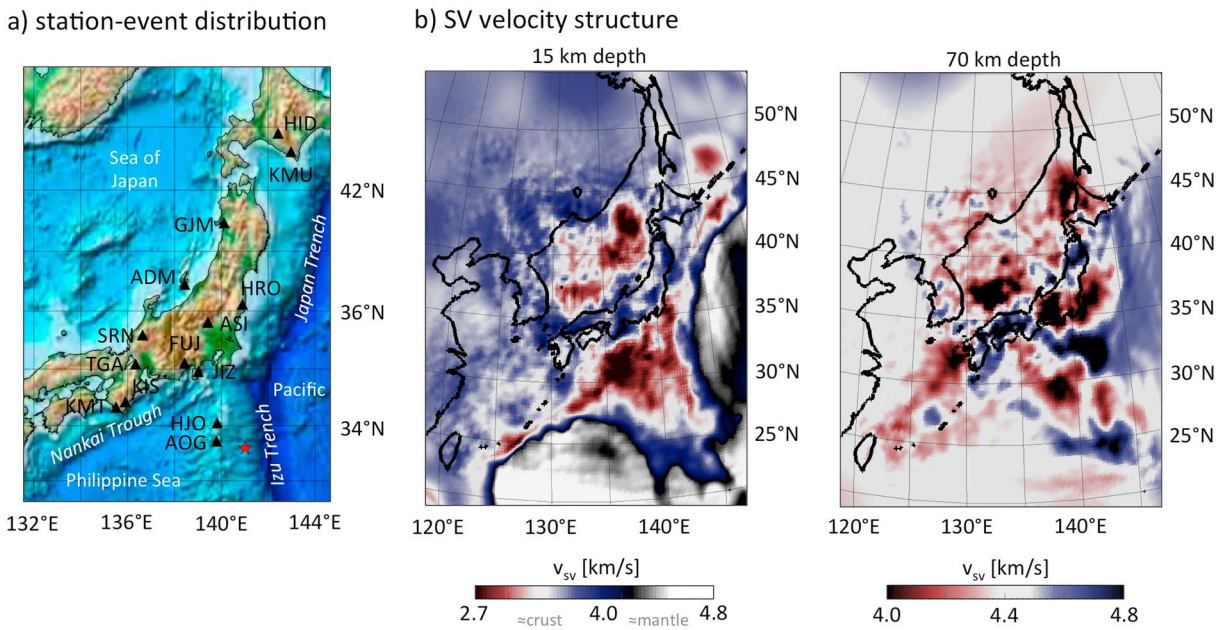

Figure 5. Source-receiver geometry and the structural model. (a) Distribution of broadband stations (black triangles) used in the source inversion. The a priori epicentral location of the earthquake is marked by a red star. (b) SV velocity structure at 15 and 70 km depth in the radially anisotropic model of the Japanese Islands used for the computation of synthetic seismograms. In the 15 km depth slice, gray scale approximately marks mantle velocities, whereas colors from red over white to blue approximately represent crustal velocities.

The most significant interparameter trade-offs are shown in Figure 8 by a collection of 2-D posterior marginals. They illustrate, for instance, that both M_{xx} and M_{yy} may be reduced without changing the misfit significantly, which is in accord with the previous finding that the trace is rather poorly constrained. To assess convergence of a 2-D trade-off measure, we compute the normalized posterior covariance

$$\mathrm{cov}_0(q_i, q_j) = \frac{\mathrm{cov}(q_i, q_j)}{\sigma_i \sigma_j}, \tag{17}$$

where σ_i denotes the posterior standard deviation of parameter q_i. Following the color scheme from Figure 7, values of cov_0 are shown in blue for 1 million samples and in red for 100 samples. Since the computation of a covariance involves the square of parameters, the differences are larger than for the single-parameter standard deviation, but still mostly limited to around 10%. The only exception is the parameter pair t_0 latitude, where larger differences result from the division by a small standard deviation in equation (17).

6. Discussion

In the previous sections, we introduced the concept of Hamiltonian Monte Carlo in the context of geophysical Bayesian inference and showed applications to seismic point source inversion for regional CMTs and location. Issues that remain to be discussed include the extent to which Hamiltonian Monte Carlo can be expected to be universally efficient, the parametrization of the moment tensor, the quantification of observational and modeling uncertainties, and the role of the prior.

6.1. No Free Lunch

While Hamiltonian Monte Carlo is widely considered to be an efficient alternative to preexisting sampling methods, it is important to keep in mind that efficiency only exists within a context. As stated by a series of No-Free-Lunch theorems (e.g., Mosegaard, 2012; Wolpert & Macready, 1997), efficiency is not an inherent attribute of a method, but it arises from the combination of the method with the right kind of prior knowledge about a specific problem.

In the case of seismic source inversion, we possess the prior knowledge that the forward problem is exactly linear in the parameters that can vary by many orders of magnitude, that is, the moment tensor components. Furthermore, we know that derivatives with respect to other parameters, including origin time and location,

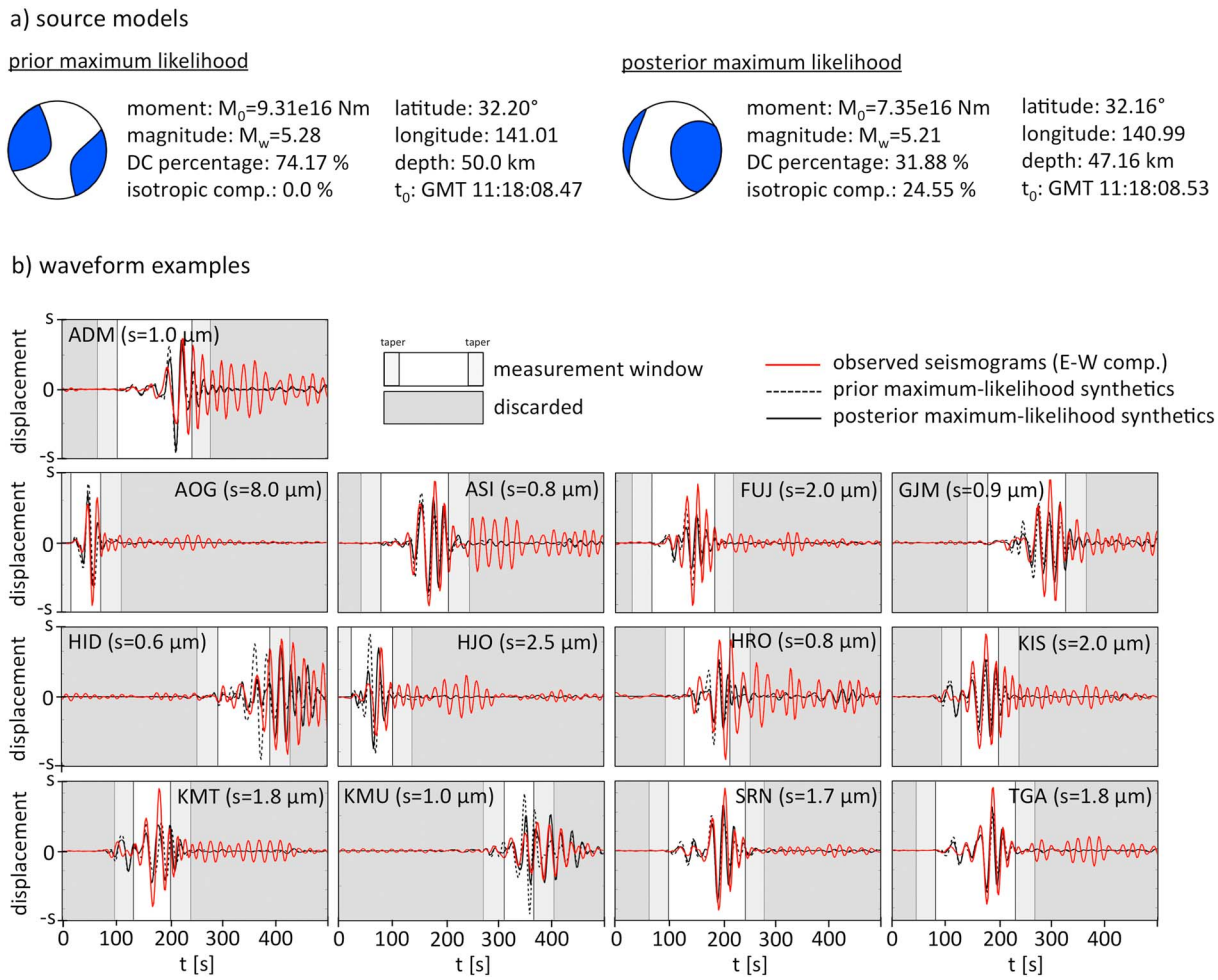

Figure 6. Summary of source models and waveforms. (a) Properties of the prior and posterior maximum-likelihood source models, the latter being obtained from 100 samples. DC percentage is the double-couple percentage as defined in Jost and Herrmann (1989). (b) E-W components seismograms. Observations are plotted in red, synthetics for the prior maximum-likelihood source in black dashed, and synthetics for the posterior maximum-likelihood source from 100 samples in black solid. The measurement windows used in the Hamiltonian Monte Carlo sampling are marked in white, with boundaries of the cosine taper in light gray. Time intervals marked in darker gray have been discarded because the initial fit between observations and synthetics was insufficient. Waveforms for the N-S and vertical components appear similar and are therefore not shown.

can be computed easily, and these derivatives enable a meaningful linearization. Without these prerequisites, Hamiltonian Monte Carlo would not be a good choice for the source inversion problem.

Along similar lines, we note that Hamiltonian Monte Carlo is not the only method designed to reduce the impact of the curse of dimensionality and to increase the model space dimension that can be handled in practical applications. Others include the Neighborhood Algorithm (Sambridge, 1999a, 1999b), (parallel) tempering (e.g., Geyer & Thompson, 1995; Marinari & Parisi, 1992; Sambridge, 2014), and transdimensional sampling (e.g., Bodin & Sambridge, 2009; Green, 1995; Sambridge et al., 2006, 2013).

6.2. Parametrization

Throughout the history of seismic source inversion methods, numerous parametrizations have been proposed, especially of the moment tensor. These include the classic decomposition into isotropic, double-couple, and compensated linear-vector components (Kikuchi & Kanamori, 1991; Knopoff & Randall, 1970); decompositions based on the eigenvectors and eigenvalues of the moment tensor (e.g., Chapman & Leaney, 2012; Hudson et al., 1989; Riedesel & Jordan, 1989); and the uniform parametrization of the moment tensor space (e.g., Tape & Tape, 2012, 2015).

All of the proposed parametrizations fully span the space of moment tensors. This turns the choice of one of them into a delicate balancing act between advantages and disadvantages that is to some extent subjective.

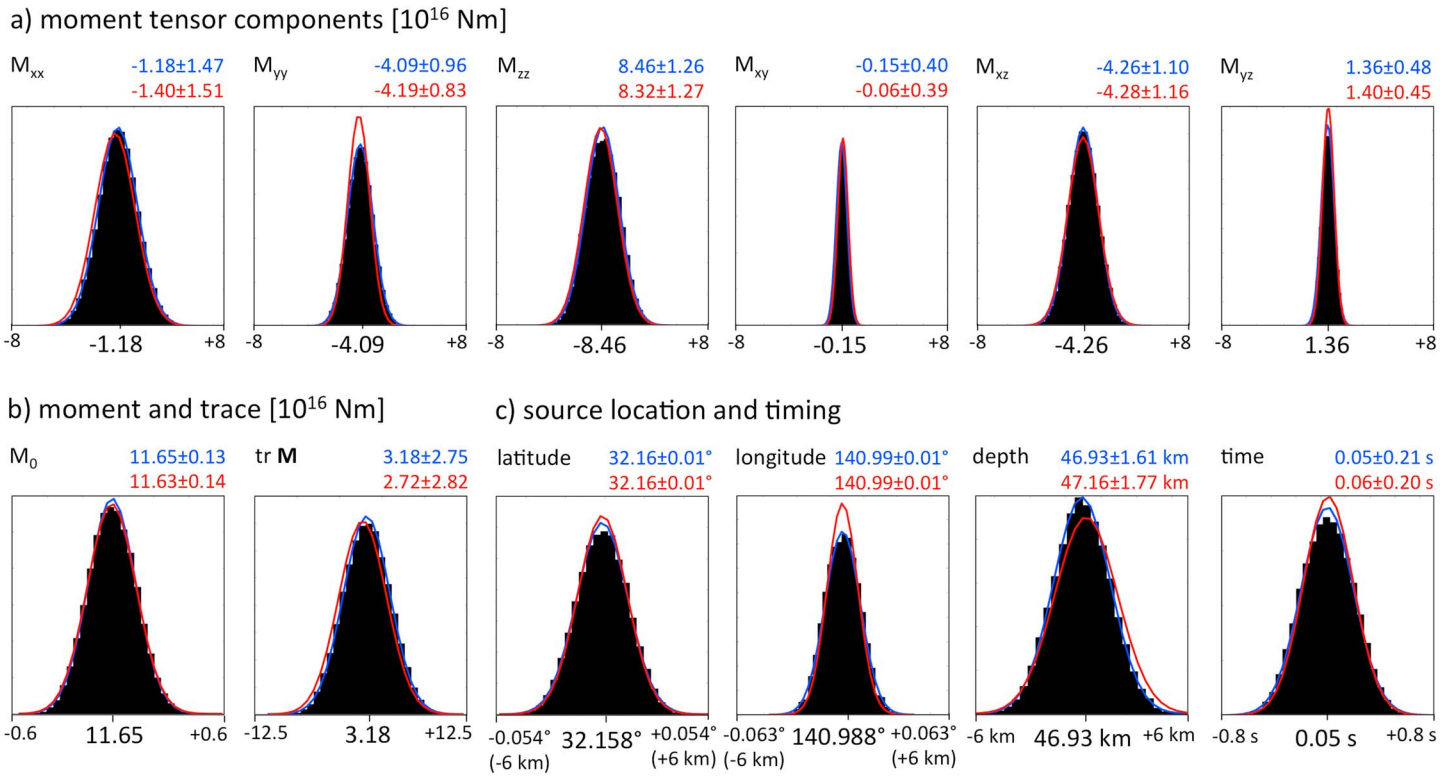

Figure 7. Summary of one-dimensional posterior marginals. (a) Marginals of the of the individual moment tensor components in units of 10^{16} Nm. Black histograms are drawn from 1 million samples. The posterior mean, standard deviation, and Gaussian are shown for 1 million samples in blue and for 100 samples in red. For better comparability, the *x* axes are labeled with the mean and deviations from the mean, for example, -1.18 ± 8 for M_{xx}. (b) Posterior distributions of the seismic moment, M_0, and the moment tensor trace, tr **M**. The color coding is the same as above; blue for 1 million samples and red for only 100 samples. (c) Posterior marginals for source location and timing, relative to the prior origin time. Color coding and labeling as before.

While, for instance, the parametrizations of Chapman and Leaney (2012), Hudson et al. (1989), Riedesel and Jordan (1989), and Tape and Tape (2015) may help with the physical interpretation of the moment tensor, they also sacrifice the linearity of the forward problem. An advantage of the uniform parameterization proposed by Tape and Tape (2015) is that a larger number of parameters is naturally bounded, which could make Metropolis-Hastings sampling more competitive.

Besides subjective preferences, Hamiltonian Monte Carlo relies on the efficient computation of derivatives, which gives a clear preference to parametrizations that lead to a linear forward problem. Since Monte Carlo sampling does not require artificial regularization by suppressing isotropic or non-double-couple components, our choice of working directly with the moment tensor components seems natural.

6.3. Quantification of Modeling and Observational Errors

With the focus of this work being on Hamiltonian Monte Carlo as a method for source inversion, we admittedly paid less attention to other aspects of the problem that may not be less important. These include the quantification of forward modeling and observational uncertainties, as performed, for instance, by Mustać and Tkalčić (2016), Silwal and Tape (2016), Staehler and Sigloch (2014, 2017), Vackár et al. (2017), and Wéber (2006).

In the interest of a clear methodological description, we limited ourselves to the assumption of uncorrelated Gaussian observational errors. Based on the analysis of preevent noise, we concluded that they are well (but still conservatively) described by a single standard deviation of $\sigma_d = 0.1\mu$m for all stations. Since the structural model used to compute synthetic seismograms is the result of a full seismic waveform inversion in the region of interest, we further assumed that forward modeling errors are negligible, relative to the observational uncertainties.

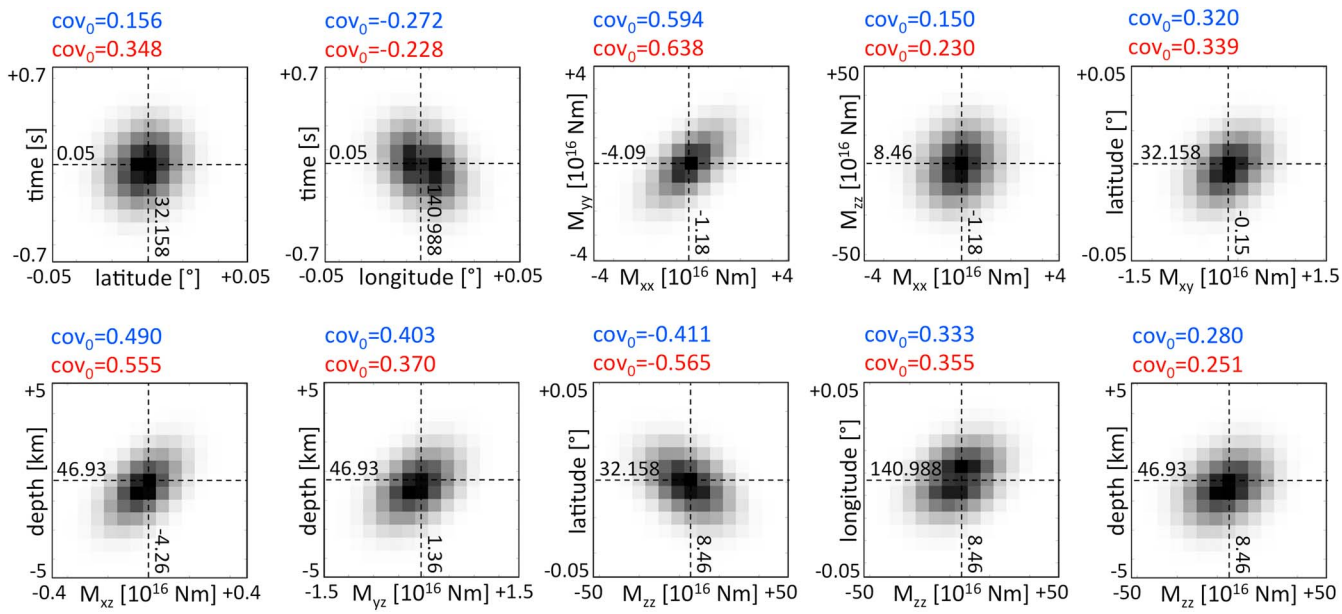

Figure 8. Summary of the most significant interparameter trade-offs in the form of 2-D posterior marginal distributions. All other parameter pairs are nearly independent. Posterior maximum-likelihood values are indicated inside the panels. Following the color coding of Figure 7, normalized covariances, $\mathrm{cov}_0(q_i, q_j) = \mathrm{cov}(q_i, q_j)/(\sigma_i \sigma_j)$, are shown in blue for 1 million samples, and in red for 100 samples.

Regardless of these pragmatic simplifications, we note that more complex descriptions of observational and forward modeling uncertainties can be incorporated into Hamiltonian Monte Carlo by including proper covariance matrices in the definition of the potential energy in equation (9).

6.4. Role of the Prior

An essential component of our "sloppy" Hamiltonian Monte Carlo is a Taylor expansion around the prior mean model. As a consequence, the quality of the prior mean — in terms of its proximity to the posterior mean — can affect the convergence of the algorithm for those parameters that do not enter the forward problem exactly linearly (e.g., source location and origin time). The effect on convergence is hard to quantify a priori and should, as so often, be assessed on a case-by-case basis. Parenthetically, we remark that the proposed method may be modified by replacing the Taylor expansion around the prior mean by a Taylor expansion around the current model.

6.5. Computational Requirements

The computational requirements of the source inversions shown in the previous sections are generally low, meaning that they are feasible on modern laptop computers. While precise numbers are too hardware dependent to be generally meaningful, we note that the CPU time for the real-data inversion is on the order of several seconds. Wavefield storage requirements are on the order of several gigabytes, though this may be reduced drastically using modern wavefield compression techniques (e.g., Boehm et al., 2016).

7. Conclusions

We present a new method for the Bayesian inference of effectively point-localized seismic sources, including their timing, location, moment tensor, and source time function. The method has two key elements: (1) Spectral element simulations combined with wavefield reciprocity allow us to account for 3-D heterogeneous Earth structure. (2) Hamiltonian Monte Carlo sampling drastically reduces the number of required samples, especially in those cases where the data quality is high and/or prior knowledge is weak.

Since Hamiltonian Monte Carlo is sill hardly used for the solution of seismological inverse problems, we provide a detailed introduction, with special emphasis on weakly nonlinear problems for which first derivatives can be computed efficiently. Seismic source inversion falls into this category, meaning that it is within the niche of applications for which Hamiltonian Monte Carlo is efficient. As part of the introduction, we show

that the acceptance rate of Hamiltonian Monte Carlo is practically unaffected when data quality increases and when prior knowledge weakens. This is in strong contrast, for instance, to the widely used Metropolis-Hastings algorithm.

In a real-data application for an event in the Japanese Islands region, we demonstrate the practical feasibility of our Hamiltonian Monte Carlo source inversion. Using a 3-D structural model inferred from full-waveform inversion, we obtain the full posterior for timing, location, and mechanism with as little as 100 samples.

While details of the method, including the treatment of observational errors and the measurement process, may still be improved, the examples are promising. They indicate that extensions to larger-scale finite-source inversions with many more parameters may be possible.

Acknowledgments
The authors gratefully acknowledge discussions with Lars Gebraad, Klaus Mosegaard, Amir Khan, Malcolm Sambridge, Peter-Jan van Leeuwen, and Andrea Zunino. Ryan Modrak and an anonymous reviewer help substantially to improve the manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 714069). Furthermore, this research was supported by the Swiss National Supercomputing Center (CSCS) in the form of the CH1 project. All seismic waveform data used in this study are freely available from the Full Range Seismograph Network of Japan (F-Net, http://www.fnet.bosai.go.jp), the Broadband Array in Taiwan for Seismology (BATS, http://bats.earth.sinica.edu.tw), the Korea National Seismograph Network (http://www.kma.go.kr/weather/earthquake/internationallist.jsp), and the China National Seismic Network, the New China Digital Seismograph Network, the Global Seismograph Network, and the Korean Seismic Network, made available by the IRIS Data Management Center (http://ds.iris.edu/ds/nodes/dmc/). All data sources were last accessed in March 2017.

References

Aki, K., & Patton, H. (1978). Determination of seismic moment tensor using surface waves. *Tectonophysics*, *49*, 213–222.
Aki, K., & Richards, P. (2002). *Quantitative seismology*. Sausalito: University Science Books.
Backus, G. E., & Mulcahy, M. (1976a). Moment tensors and other phenomenological descriptions of seismic sources—I. Continuous displacements. *Journal of the Royal Astronomical Society*, *46*, 341–361.
Backus, G. E., & Mulcahy, M. (1976b). Moment tensors and other phenomenological descriptions of seismic sources—I. Continuous displacements. *Journal of the Royal Astronomical Society*, *46*, 341–361.
Bernardi, F., Braunmiller, J., Kradolfer, U., & Giardini, D. (2004). Automatic regional moment tensor inversion in the European-Mediterranean region. *Geophysical Journal International*, *157*, 703–716.
Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [stat.ME]*.
Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, *178*, 1411–1436.
Boehm, C., Hanzich, M., de la Puente, J., & Fichtner, A. (2016). Wavefield compression for adjoint methods in full-waveform inversion. *Geophysics*, *81*, R385–R397.
Buland, R., & Gilbert, F. (1976). Matched filtering for the seismic moment tensor. *Geophysical Research Letters*, *3*, 205–206.
Burridge, R., & Knopoff, L. (1964). Body force equivalents for seismic dislocations. *Bulletin of the Seismological Society of America*, *54*, 1875–1888.
Cesca, S., Heimann, S., Stammler, K., & Dahm, T. (2010). Automated procedure for point and kinematic source inversion at regional distances. *Journal of Geophysical Research*, *115*, B06304. https://doi.org/doi:10.1029/2009JB006450
Chapman, C. H., & Leaney, W. S. (2012). A new moment-tensor decomposition for seismic events in anisotropic media. *Geophysical Journal International*, *188*, 343–370.
Chen, P., Zhao, L., & Jordan, T. H. (2007). Full 3D tomography for the crustal structure of the Los Angeles region.*Bulletin of the Seismological Society of America*, *97*, 1094–1120.
Donner, S., Bernauer, M., & Igel, H. (2016). Inversion for seismic moment tensors combining translational and rotational ground motions. *Geophysical Journal International*, *207*, 562–570.
Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, *195*, 216–222.
Dubbledam, D., Calero, S., Ellis, D. E., & Snurr, R. Q. (2016). RASPA: Molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Molecular Simulation*, *42*, 81–101.
Dufumier, H., & Rivera, L. (1997). On the resolution of the isotropic component in moment tensor inversion. *Geophysical Journal International*, *131*, 595–606.
Dziewoński, A. M., & Gilbert, F. (1974). Temporal variation of the seismic moment tensor and the evidence for precursive compression for two deep earthquakes. *Nature*, *247*, 185–188.
Dziewoński, A. M., & Woodhouse, J. H. (1981). An experiment in systematic study of global seismicity: Centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981. *Journal of Geophysical Research*, *88*, 3247–3271.
Dziewoński, A. M., Chou, T.-A., & Woodhouse, J. H. (1981). Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *Journal of Geophysical Research*, *10*, 2825–2852.
Elhatisari, S., Lee, D., Rupak, G., Epelbaum, E., Krebs, H., Lahde, T. A., et al. (2015). Ab initio alpha-alpha scattering. *Nature*, *258*, 111–113.
Faccioli, E., Maggio, F., Paolucci, R., & Quarteroni, A. (1997). 2D and 3D elastic wave propagation by a pseudospectral domain decomposition method. *Journal of Seismology*, *1*, 237–251.
Fichtner, A. (2010). *Full seismic waveform modelling and inversion*. Heidelberg: Springer.
Fichtner, A., & Tkalcic, H. (2010). Insights into the kinematics of a volcanic caldera drop: Probabilistic finite-source inversion of the 1996 Bardarbunga, Iceland, earthquake. *Earth and Planetary Science Letters*, *297*, 607–615.
Fichtner, A., Bunge, H.-P., & Igel, H. (2006). The adjoint method in seismology—I. Theory. *Physics of the Earth and Planetary Interiors*, *157*, 86–104.
Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H.-P. (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. *Geophysical Journal International*, *175*, 665–685.
Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H.-P. (2009). Spectral-element simulation and inversion of seismic waves in a spherical section of the Earth. *Japan Journal of Industrial and Applied Mathematics*, *4*, 11–22.
Geyer, C. J., & Thompson, E. A. (1995). Annealing Markov Chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, *90*, 909–920.
Gilbert, F. (1973). Derivation of source parameters from low-frequency spectra. *Philosophical Transactions of the Royal Society A*, *274*, 369–371.
Gokhberg, A., & Fichtner, A. (2016). Full-waveform inversion on heterogeneous HPC systems. *Computers & Geosciences*, *89*, 260–268.
Green, P. J. (1995). Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732.
Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, *57*, 97–109.

Hejrani, B., Tkalcic, H., & Fichtner, A. (2017). Centroid moment tensor catalogue using a 3D continental scale Earth model: Application to earthquakes in Papua New Guinea and the Solomon Islands. *Journal of Geophysical Research: Solid Earth*, *122*, 5517–5543. https://doi.org/10.1002/2017JB014230

Hingee, M., Tkalčić, H., Fichtner, A., & Sambridge, M. (2011). Seismic moment tensor inversion using a 3-D structural model: Applications for the Australian region. *Geophysical Journal International*, *184*, 949–964.

Hoffmann, M. D., & Gelman, A. (2014). The No-U-Turn sampler. *Journal of Machine Learning Research*, *15*, 1593–1623.

Honkela, A., Peltonen, J., Topa, H., Charapitsa, I., Matarese, F., Grote, K., et al. (2015). Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 13,115–13,120.

Hsieh, M., Zhao, L., & Ma, K. (2014). Efficient waveform inversion for average earthquake rupture in three-dimensional structures. *Geophysical Journal International*, *198*, 1279–1292.

Hudson, J. A., Pearce, R. G., & Rogers, R. M. (1989). Source type plot for inversion of the moment tensor. *Journal of Geophysical Research*, *94*, 765–774.

Jost, M. L., & Herrmann, R. B. (1989). A student's guide to and review of moment tensors. *Seismological Research Letters*, *60*, 37–57.

Käufl, P., Valentine, A., de Wit, R., & Trampert, J. (2015). Robust and fast probabilistic source parameter estimation from near-field displacement waveforms using pattern recognition. *Bulletin of the Seismological Society of America*, *105*, 2299–2312. https://doi.org/10.1785/0120150010

Kikuchi, M., & Kanamori, H. (1991). Inversion of complex body waves—III. *Bulletin of the Seismological Society of America*, *81*, 2335–2350.

Knopoff, L., & Randall, M. J. (1970). The compensated linear-vector dipole: A possible mechanism for deep earthquakes. *Journal of Geophysical Research*, *54*, 4957–4963.

Komatitsch, D., & Vilotte, J. P. (1998). The spectral element method: An effective tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America*, *88*, 368–392.

Krischer, L., Fichtner, A., Žukauskaitė, S., & Igel, H. (2015). Large-scale seismic inversion framework. *Seismological Research Letters*, *86*, 1198–1207.

Landau, L. D., & Lifshitz, E. M. (1976). *Course of theoretical physics, Volume 1, Mechanics* (3rd ed.). Amsterdam: Elsevier Butterworth Heinemann.

Lee, E.-J., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M., & Beroza, G. C. (2014). Full-3D tomography (F3DT) for crustal structure in Southern California based on the scattering-integral (SI) and the adjoint-wavefield (AW) methods. *Journal of Geophysical Research: Solid Earth*, *119*, 6421–6451. https://doi.org/10.1002/2014JB011346

Liu, Q., Polet, J., Komatitsch, D., & Tromp, J. (2004). Spectral-element moment tensor inversion for earthquakes in Southern California. *Bulletin of the Seismological Society of America*, *94*, 1748–1761.

Marinari, E., & Parisi, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters*, *19*, 451–458.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*, 1087–1092.

Mosegaard, K. (2012). *Limits to nonlinear inversion* (pp. 11–21). Berlin, Heidelberg: Springer.

Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, *100*, 12,431–12,447.

Muir, J. B., & Tkalčić, H. (2015). Probabilistic joint inversion of lowermost mantle P-wave velocities and core mantle boundary topography using differential travel times and hierarchical Hamiltonian Monte-Carlo sampling. In *AGU 2015 Fall meeting*. S14A–03, AGU.

Mustać, M., & Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, *204*, 311–323.

Neal, R. M. (1996). *Bayesian learning for neural networks*. New York: Springer.

Neal, R. M. (2011). MCMV using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* (Chap 5).

O'Toole, T. B., Valentine, A. P., & Woodhouse, J. H. (2012). Centroid-moment tensor inversions using high-rate GPS waveforms. *Geophysical Journal International*, *191*, 257–270.

Randall, M. J., & Knopoff, L. (1970). The mechanism at the focus of deep earthquakes. *Journal of Geophysical Research*, *75*, 4965–4976.

Riedesel, M. A., & Jordan, T. H. (1989). Display and assessment of seismic moment tensors. *Bulletin of the Seismological Society of America*, *79*, 85–100.

Sambridge, M. S. (1999a). Geophysical inversion with the Neighbourhood Algorithm—I. Searching a parameter space. *Geophysical Journal International*, *138*, 479–494.

Sambridge, M. S. (1999b). Geophysical inversion with the Neighbourhood Algorithm—II. Appraising the ensemble. *Geophysical Journal International*, *138*, 727–746.

Sambridge, M. S. (2014). A parallel tempering algorithm for probabilistic sampling and multi-modal optimization. *Geophysical Journal International*, *196*, 357–374.

Sambridge, M. S., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, *40*(3), 1009. https://doi.org/doi:10.1029/2000RG000089

Sambridge, M. S., Bodin, T., Gallagher, K., & Tkalcic, H. (2013). Transdimensional inference in the geosciences. *Philosophical Transactions of the Royal Society A*, *371*(1984), 20110547. https://doi.org/doi: 10.1098/rsta.2011.0547

Sambridge, M. S., Gallagher, K., Jackson, A., & Rickwood, P. (2006). Trans-dimensional inverse problems, model comparison, and the evidence. *Geophysical Journal International*, *167*, 528–542.

Scognamiglio, L., Tinti, E., & Michelini, A. (2009). Real-time determination of seismic moment tensor for the Italian region. *Bulletin of the Seismological Society of America*, *99*, 2223–2242.

Seah, Y.-L., Shang, J., Ng, H. K., Nott, D. J., & Englert, B.-G. (2015). Monte Carlo sampling from the quantum state space. II. *New Journal of Physics*, *17*, 043017. https://doi.org/doi:10.1088/1367–2630/17/4/043018

Sen, M. K., & Biswas, R. (2017). Tansdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm. *Geophysics*, *82*, R119–R134.

Silwal, V., & Tape, C. (2016). Seismic moment tensors and estimated uncertainties in southern Alaska. *Journal of Geophysical Research: Solid Earth*, *121*, 2772–2797. https://doi.org/10.1002/2015JB012588

Simute, S., Steptoe, H., Gokhberg, A., & Fichtner, A. (2016). Full-waveform inversion of the Japanese Islands region. *Journal of Geophysical Research: Solid Earth*, *121*, 3722–3741. https://doi.org/10.1002/2016JB012802

Staehler, S. C., & Sigloch, K. (2014). Fully probabilistic seismic source inversion—Part I: Efficient parameterisation. *Solid Earth*, *5*, 1055–1069. https://doi.org/doi:10.5194/se–5–1055–2014

Staehler, S. C., & Sigloch, K. (2017). Fully probabilistic seismic source inversion—Part II: Modelling errors and station covariances. *Solid Earth*, *7*(6), 1521–1536. https://doi.org/doi:10.5194/se–7–1521–2016

Symon, K. R. (1971). *Mechanics, Addison-Wesley*. Reading: Massachusetts.
Tape, W., & Tape, C. (2012). A geometric setting for moment tensors. *Geophysical Journal International*, *190*, 476–498.
Tape, W., & Tape, C. (2015). A uniform parametrization of moment tensors. *Geophysical Journal International*, *202*, 2074–2081.
Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. *Pure and Applied Geophysics*, *128*, 365–399.
Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation* (2nd ed.). Philadephia: Society for Industrial and Applied Mathematics.
Tromp, J., Tape, C., & Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, *160*, 195–216.
Vackár, J., Burjének, J., Gallovič, F., Zahradnik, J., & Clinton, J. (2017). Bayesian ISOLA: New tool for automated centroid moment tensor inversion. *Geophysical Journal International*, *210*, 693–705.
Valentine, A. P., & Trampert, J. (2012). Assessing the uncertainties on seismic source parameters: Towards realistic error estimates for centroid moment-tensor determinations. *Physics of the Earth and Planetary Interiors*, *210-211*, 36–49.
Vallée, M. (2013). Source time function properties indicate a strain drop independent of earthquake depth and magnitude. *Nature Communications*, *4*, 2606. https://doi.org/doi:10.1038/ncomms3606
Wéber, Z. (2006). Probabilistic local waveform inversion for moment tensor and hypocentral location. *Geophysical Journal International*, *165*, 607–621.
Wolpert, D. H., & Macready, W. G. (1997). No Free Lunch Theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*, 67–82.
Zhao, L., Chen, P., & Jordan, T. H. (2006). Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies. *Bulletin of the Seismological Society of America*, *96*, 1753–1763.
Zhu, L., & Zhou, X. (2016). Seismic moment tensor inversion using 3D velocity model and its application to the 2013 Lushan earthquake sequence. *Physics and Chemistry of the Earth*, *95*, 10–18. https://doi.org/doi: 10.1016/j.pce.2016.01.002