Bayesian seismic source inversion with a 3-D Earth model of the Japanese Islands

Saule Simute[1], Christian Boehm[2], Lion Krischer[3], Alexey Gokhberg[4], Martin Vallée[5], and Andreas Fichtner[1]

[1]ETH Zurich
[2]Department of Earth Sciences, Institute of Geophysics, ETH Zürich
[3]Mondaic AG
[4]Fragata Computer Systems AG
[5]Institut De Physique Du Globe De Paris

November 21, 2022

Abstract

We present probabilistic centroid-moment tensor solutions inferred from the combination of Hamiltonian Monte Carlo sampling and a 3-D full-waveform inversion Earth model of the Japanese archipelago. While the former provides complete posterior probability densities, the latter allows us to exploit waveform data with periods as low as 15 s. For the computation of Green's functions, we employ spectral-element simulations through the radially anisotropic and visco-elastic model, leading to substantial improvements of data fit compared to layered models. Focusing on Mw 4.8 - Mw 5.3 offshore earthquakes with a significant non-double-couple component, we simultaneously infer the centroid location, time and moment tensor without any a priori constraints on the faulting mechanism. Furthermore, we perform the inversions across several period bands, varying the minimum period between 15 s and 50 s. Accounting for 3-D Earth structure at shorter periods can increase the double-couple component of an event, compared to the GCMT solution, by tens of percent. This suggests that at least some of the non-double-couple events in the GCMT catalog might result from unmodeled Earth structure. We also observe that significant changes in source parameters, and the double-couple component in particular, may be related to only small waveform changes, thereby accentuating the importance of a reliable Earth model. Posterior probability density distributions become increasingly multimodal for shorter-period data that provide tighter constraints on source parameters. This implies, in our specific case, that stochastic approaches to the source inversion problem are required for periods below ~ 20 s to avoid trapping in local minima.

Bayesian seismic source inversion with a 3-D Earth model of the Japanese Islands

Saulė Simutė[1,2], Christian Boehm [1], Lion Krischer [1], Alexey Gokhberg [1,3], Martin Vallée [2], Andreas Fichtner [1]

[1]Institute of Geophysics, ETH Zurich, Zurich, Switzerland
[2]Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France
[3]Fragata Computer Systems AG, Schwyz, Switzerland

Key Points:

- Full-waveform inversion models enable the use of shorter-period data in earthquake source inversion.
- Using regional waveforms at 15 s period may reduce non-double-couple components by tens of percent.
- Probability densities of source parameters below 20 s are non-Gaussian, thus demanding stochastic inversion approaches.

Corresponding author: Saulė Simutė, `saaule@gmail.com`

Abstract

We present probabilistic centroid-moment tensor solutions inferred from the combination of Hamiltonian Monte Carlo sampling and a 3-D full-waveform inversion Earth model of the Japanese archipelago. While the former provides complete posterior probability densities, the latter allows us to exploit waveform data with periods as low as 15 s. For the computation of Green's functions, we employ spectral-element simulations through the radially anisotropic and visco-elastic model, leading to substantial improvements of data fit compared to layered models. Focusing on M_w 4.8 – M_w 5.3 offshore earthquakes with a significant non-double-couple component, we simultaneously infer the centroid location, time and moment tensor without any a priori constraints on the faulting mechanism. Furthermore, we perform the inversions across several period bands, varying the minimum period between 15 s and 50 s. Accounting for 3-D Earth structure at shorter periods can increase the double-couple component of an event, compared to the GCMT solution, by tens of percent. This suggests that at least some of the non-double-couple events in the GCMT catalog might result from unmodeled Earth structure. We also observe that significant changes in source parameters, and the double-couple component in particular, may be related to only small waveform changes, thereby accentuating the importance of a reliable Earth model. Posterior probability density distributions become increasingly multimodal for shorter-period data that provide tighter constraints on source parameters. This implies, in our specific case, that stochastic approaches to the source inversion problem are required for periods below ~ 20 s to avoid trapping in local minima.

Plain Language Summary

In the majority of global earthquake catalogs, the earthquake solution, i.e., centroid location, time and a rupture mechanism, is typically inferred assuming a 1-D Earth model. However, both earthquake source and Earth structure contribute to seismic recordings, meaning that unaccounted structure might map into and pollute the source solution. In this study we use a 3-D Earth structure of the Japanese archipelago to model the waveforms and infer earthquake parameters of small-to-moderate magnitude offshore events. We do not put any a priori constraints on the faulting mechanism and let it be determined by the data. We perform stochastic inversions which provide us with a collection of all plausible models ranked by their respective probability. When a 3-D Earth structure at shorter periods is taken into account, the earthquake mechanism can be largely explained by a slip on the fault. We also observe that significant changes in source parameters may be related to tiny waveform changes, thereby accentuating the importance of a reliable Earth model.

1 Introduction

Earthquake source solutions are important in many fields of seismology, such as, but not limited to, seismic hazard, earthquake physics, seismotectonics, and seismic tomography. Although source inversion is an established discipline in seismology, obtaining a robust solution of moment tensor components together with a spatial and temporal location is still a challenging task. A largely unexplored potential lies in the adoption of more realistic Green's functions, which until recently have only accounted for radially symmetric Earth structure. In the wake of increasing computational power and growing number of full-waveform tomographic models, use of numerically computed Green's functions for complex regional Earth models has become possible.

1.1 Recent developments in source inversion

Earthquake source mechanisms, in terms of the first-motion polarities and fault-plane solutions, have been studied since the beginning of the 20th century (Omori, 1905; Galitzin, 1909; Byerly, 1928), and the first computer programs, intended to aid the graphical analysis, were developed in the early 1960s (Knopoff, 1961; Kasahara, 1963). A significant development was accomplished by Backus and Mulcahy (1976a, 1976b), who derived a phenomenological representation for an indigenous source and showed that each seismic source can be described by a moment tensor, or a distribution thereof. Seismic source inversion has become routine since the end of the last century (e.g., Mendiguren, 1977; Kanamori & Given, 1981; Dziewoński et al., 1981). Since then, different approaches have been established to retrieve information on the source parameters, based on, e.g., first-motion polarity (e.g., Knopoff, 1961; Kasahara, 1963; Lentas, 2017; Hara et al., 2019), body waveforms (e.g., Dziewoński et al., 1981; Dreger & Helmberger, 1991; Vallée et al., 2011), surface waves (e.g., Kanamori & Given, 1981; Romanowicz, 1982; Ferreira & Woodhouse, 2006), with a specific interest to the ultra-long-period W-phase (e.g., Kanamori & Rivera, 2008; Duputel et al., 2012; Hayes et al., 2009), or full waveforms, incorporating both body and surface wave signals (e.g., Dreger, 2003; Ekström et al., 2012; Scognamiglio et al., 2016; Hallo et al., 2017). While some methods might be more robust than others, all of them, to some degree, rely on how well one can predict the data for a given set of model parameters. The choice of the Earth model, hence, is of fundamental importance in earthquake source inversion, as unaccounted Earth structure might map into the source solution and potentially pollute it (e.g., Hjörleifsdóttir & Ekström, 2010; Woodhouse, 1983; Smith & Ekström, 1996; Thurber, 1983).

Until recently, radially symmetric Earth models have been predominantly used in source inversion studies, for they allow one to efficiently compute Green's functions. Lateral heterogeneities are then taken into account via empirical or theoretical corrections (Ferreira et al., 2011). For example, traces or even the different portions of Green's functions can be shifted independently to fit the data (e.g., Zhao & Helmberger, 1994; Zhu & Helmberger, 1996; Ford et al., 2009a, 2009b). However, such corrections might mask earthquake source effects. Theoretical surface wave corrections may be implemented in terms of mean phase slowness along the source-receiver great circle (e.g., Woodhouse & Dziewoński, 1984; Pondrelli et al., 2002), neglecting the amplitude effects. In addition, great circle approximations, relying on ray theory, do not account for finite-frequency effects of wave propagation, hence the corrections themselves might be erroneous. Another approach is to use multiple 1-D Earth models to account for differences in oceanic and continental crust (Lee et al., 2011), as is done for the National Research Institute for Earth Science and Disaster Prevention (NIED) earthquake catalog in Japan (Kubo et al., 2002). Alternatively, the dependence on structural models can be alleviated, focusing on those data which are less sensitive to crustal heterogeneities, such as the W-phase (Kanamori & Rivera, 2008) or the Pnl phase (Helmberger & Enge, 1980).

With increasing computational power, improving numerical methods (e.g., Nissen-Meyer et al., 2007; Komatitsch et al., 2010; Krischer et al., 2015; Gokhberg & Fichtner, 2016; Afanasiev et al., 2018) and theoretical developments (e.g., Tromp et al., 2005; Fichtner, van Herwaarden, et al., 2018; Thrastarson et al., 2020; van Herwaarden et al., 2020), full-waveform tomographic models have been proliferating on both regional (e.g., Fichtner et al., 2009a; Krischer et al., 2018; Blom et al., 2020) and global scale (e.g., Bozdağ et al., 2016; French & Romanowicz, 2014; Fichtner, van Herwaarden, et al., 2018). This has in turn enabled researches to start using numerically computed 3-D Green's functions for source inversion. Such type of studies have been performed for the Southern California region (Liu et al., 2004; Zhao et al., 2006; Lee et al., 2011; X. Wang & Zhan, 2019), the Australian region (Hingee et al., 2011; Hejrani et al., 2017), the Sichuan province in China (Zhu & Zhou, 2016) and more recently for offshore earthquakes along the Nankai trough in Japan (Takemura et al., 2018, 2020).

The non-linear relationship between data and model parameters, such as centroid location and centroid time of an earthquake, make it difficult to tackle the source inversion with deterministic approaches. The least-squares method, for example, provides a single solution and does not account for non-uniqueness, which can arise due to insufficient data coverage and modeling inaccuracies. Furthermore, uncertainty information, derived by linearization methods, is only representative if the objective functional is indeed quadratic or otherwise have little meaning at all (Sambridge & Mosegaard, 2002). To tackle these challenges, we resort to probabilistic inference, which provides a collection of all plausible models ranked by their respective probability. Statistical inferences can be made from the ensemble to assess the uncertainty, and covariance matrices can be recovered to study the inter-parameter trade-offs. Such an approach respects the non-uniqueness, avoids the subjective regularization required by the deterministic inversion, and delivers uncertainty measures as part of the solution. However, a more vigorous exploration of the model space typically comes with higher computational costs.

Stochastic approaches were used in the inversions of microseismic events (e.g., Pugh et al., 2016; Shang & Tkalčić, 2020), of events with anomalously high non-double-couple component (e.g., Mustać & Tkalčić, 2016), finite-fault inversions (e.g., Minson et al., 2014; Duputel et al., 2014; Dettmer et al., 2014), and for earthquake early warning purposes (Cua & Heaton, 2007). Only a few probabilistic studies have been performed using fully heterogeneous Earth models (e.g., Lee et al., 2011).

1.2 Hamiltonian Monte Carlo

The performance of traditional stochastic random walk methods, such as Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970), tends to scale poorly with increasing dimension (Betancourt, 2017). One way to guarantee the efficacy of sampling is through informed proposals, a strategy to ensure that the transitions largely follow the contours of high probability mass (Neal, 1996, 2011; Betancourt, 2017). Informed algorithms, such as Hamiltonian Monte Carlo (HMC) are designed to make use of the information outside of a simple target distribution evaluation at a given point (e.g., Khoshkholgh et al., 2020; Zanella, 2020). HMC relies on the gradient information of the misfit in order to guide the sampler towards the areas of high-probability mass. It can be regarded as a hybrid approach encompassing the virtues of both gradient-based optimization and derivative-free Markov chain Monte Carlo methods (Fichtner, Zunino, & Gebraad, 2018).

Hamiltonian Monte Carlo is particularly useful for multi-dimensional problems with high quality data or weakly constrained priors, which, in traditional, derivative-free sampling algorithms would result in a low acceptance rate and a slow convergence. Although introduced in the 1980s (Duane et al., 1987), HMC has only recently gained popularity in geophysics. Maiti and Tiwari (2009) implemented HMC-based neural networks to analyze well log data, Muir and Tkalčić (2020) applied HMC for a lowermost mantle study, Sen and Biswas (2017) and Biswas and Sen (2017) used HMC in 1-D and 2-D seismic inversions, respectively, while Fichtner, Zunino, and Gebraad (2018) and Gebraad et al. (2020) further proved the potential of HMC for nonlinear seismic tomography problems. Very recently Aleardi et al. (2020) used HMC in the context of dispersion curves inversion, while Koch et al. (2020) implemented adjoint HMC in the context of engineering. A variant of HMC that tunes itself while sampling was presented by Fichtner et al. (2021).

The potential of HMC in earthquake source inversion was demonstrated by Fichtner and Simutė (2018), where the HMC was adapted for efficient source studies in complex media, with synthetic examples and a real-data illustration. In this study, we largely rely on the methodology presented in Fichtner and Simutė (2018) and perform multiple source inversions with an expanded and improved data set.

1.3 Motivation and outline

Green's functions computed for laterally averaged structure are not adequate for tectonically complex areas, especially subduction zones, which require a proper incorporation of 3-D Earth structure (e.g., Engdahl et al., 1977; Igel et al., 2002). Simplified Earth models affect the estimation of the centroid location and time (e.g., Dziewoński & Woodhouse, 1983; Hjörleifsdóttir & Ekström, 2010; Ferreira & Woodhouse, 2006; Smith & Ekström, 1996; Thurber, 1983; Morales-Yáñez et al., 2020), the seismic moment (e.g., Patton & Randall, 2002), as well as the moment tensor itself (e.g., Woodhouse, 1983; Newrkla et al., 2019; Hejrani et al., 2017; Scognamiglio et al., 2016; Ferreira & Woodhouse, 2006), which often manifest as spurious non-double couple components (Zahradník et al., 2015). However, radially symmetric Earth models, which allow for a computationally efficient way to obtain Green's functions, are still commonly used in source inversion studies on the grounds that a suitable data selection might isolate data pertaining principally to the source (e.g., Woodhouse, 1983; Mustać & Tkalčić, 2016; Ford et al., 2009a; Staehler & Sigloch, 2014).

Motivated by the effects that unacounted Earth structure potentially has on earthquake source solutions, and endorsing the need for uncertainty information, we propose a stochastic earthquake source inversion, based on the Hamiltonian Monte Carlo sampling algorithm. We start by introducing a heterogeneous, viscoelastic and radially anisotropic Earth model of the crust and upper mantle beneath the Japanese islands region, which is constructed for this study (section 2). We then introduce the formulation and practical aspects of forward and inverse problems (sections 3 – 5). Finally, we present multi-period centroid moment tensor inversion results of earthquakes at the Izu-Bonin trench (section 6). Owing to the Bayesian framework, we retrieve the uncertainty information as well as the inter-parameter trade-offs. We discuss the implications as well as the limitations of the study in section 7 and draw the concluding remarks in section 8.

2 Velocity model for the Japanese islands

To reduce the effect of 3-D Earth structure on estimated source parameters, we construct a full-waveform inversion model for the Japanese islands region, building on the velocity model previously constructed by Simutė et al. (2016) on the basis of waveform data in the 20–80 s period range. The model is viscoelastic, radially anisotropic and 3-D heterogeneous. For forward and adjoint modeling, we employ the GPU-accelerated spectral-element wave equation solver SES3D (Fichtner et al., 2009b; Gokhberg & Fichtner, 2016). We use time-frequency phase misfits (Fichtner & Igel, 2008) to quantify differences between observed and synthetic waveforms within automatically selected measurement time windows where waveform similarity is sufficient to avoid cycle skips (Krischer et al., 2015). The final model is the result of an iterative conjugate-gradient minimization of the misfit, with gradients computed by adjoint techniques (Tarantola, 1988; Tromp et al., 2005; Fichtner et al., 2006). We invert for isotropic P velocity v_p, SV velocity v_{sv}, SH velocity v_{sh}, and density ρ. Furthermore, we implement viscoelastic attenuation by using the QL6 attenuation model of Durek and Ekström (1996), which is, however, kept constant throughout the inversion. Since the focus of this work is on source inversion, we refer to Simutė et al. (2016) for a more detailed and technical description of the well-established full-waveform inversion method.

Starting with the model presented in Simutė et al. (2016), we performed 14 additional iterations using waveform data with a slightly broadened period range of 15 – 80 s. First, we completed seven iterations for the larger model domain shown in Fig. 1, which we also used previously (Simutė et al., 2016). Subsequently, we performed the remaining seven iterations for a smaller domain and with additional regional data, as indicated in Fig. 1. This was intended to specifically improve that part of the model which we later use for the computation of Green's functions, needed for the Bayesian source inversion.

With this concrete application in mind, we primarily focus on waveform fit, limiting the presentation of the structural model to a short paragraph at the end of this section.

The overall waveform misfit decreased by 24 % after the first seven iterations in the larger initial domain, and by another 21 % during the subsequent seven iterations in the smaller domain. More details on the misfit evolution are shown in the supplementary Fig. S1.

In Fig. 2, we present a small but representative collection of waveform comparisons across the model domain for four M_w 5.0 – M_w 5.8 shallow- to intermediate-depth events, situated 1) at the Izu-Bonin trench, 2) off Kuyshu, 3) in the Sea of Japan, and 4) off the east coast of Honshu. Together with the selected stations they represent model parts which are relatively well covered by the data. Still, the waveform fit is not uniform across the model domain. Observed waveforms for some paths, such as between event 19 and station BO.KSK, or event 30 and the stations in central Japan, are well explained in terms of both phase and amplitude. Other paths, in contrast, are characterized by a good match in phase but show discrepancies in amplitude; for example, the path beween event 16 and station BO.ABU. The latter is a general feature observed across the majority of the traces, suggesting that the source mechanism may need improvement.

To assess the importance of using a 3-D model for source inversion in our study region, we compare computed waveforms for our 3-D model (red solid waveforms in Fig. 2) and its laterally averaged 1-D version (red dashed waveforms in Fig. 2). The 3-D full-waveform inversion model produces a substantially better waveform fit than the 1-D model, for which time shifts can be on the order of tens of seconds. For the whole-Earth 1-D model AK135 (Kennett et al., 1995) results are similar, thus corroborating that the lateral heterogeneities in our velocity model are indeed required to fit waveform data at periods between 15 – 80 s, as studied here.

In Fig. 3 we compare the whole-seismogram waveform fit at short and long periods. The root-mean square error is computed as:

$$\chi_i = \frac{\sqrt{\int_0^T [u_i^0(t) - u_i(t)]^2 dt}}{\sqrt{\int_0^T u_i^0(t)^2 dt}}, \tag{1}$$

where $u_i^0(t)$ denotes the i-component of the observations, $u_i(t)$ the i-component of the synthetic seismograms, and T is the duration of the time series. Misfit at long-periods (50–80 s) is low throughout the domain, with remaining discrepancies close to the expected noise level. Misfits at short-period (15–80 s), on the other hand, have more variability, largely correlating with the geological complexity along the source-receiver path. The implications for the source inversion will be further discussed in section 5.4.

A collection of depth and cross-sectional slices through the tomographic model in terms of deviation of the isotropic S velocity v_s from the lateral average $\bar{v}_s$ is shown in Fig. 4. We compute isotropic S velocity as $v_s = \sqrt{\frac{2}{3}v_{sv}^2 + \frac{1}{3}v_{sh}^2}$ (e.g., Babuška & Cara, 1991; Panning & Romanowicz, 2006). The lateral average $\bar{v}_s$ (Fig. S2), more depth slices (Fig. S3), and depth profiles (Fig. S4) as well as anisotropy (Fig. S5), are presented as supplementary information.

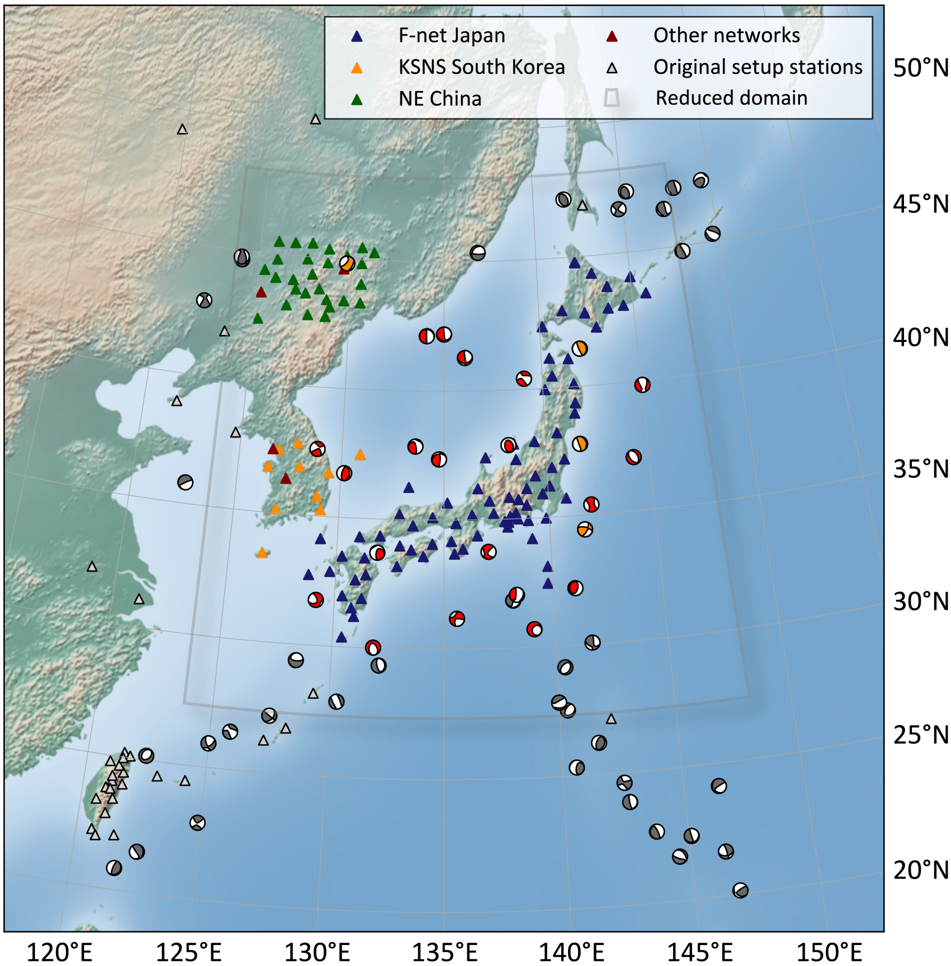

Figure 1. Source-receiver setups for tomographic inversions in the initial large domain and the smaller focused domain. Within the large domain, we used 58 earthquakes, depicted as red and grey focal mechanisms, and all the stations except for the NE China array, shown as triangles. The smaller domain comprises 20 events from the original setup shown in red and four new events in orange together with all seismic stations depicted in non-grey color.

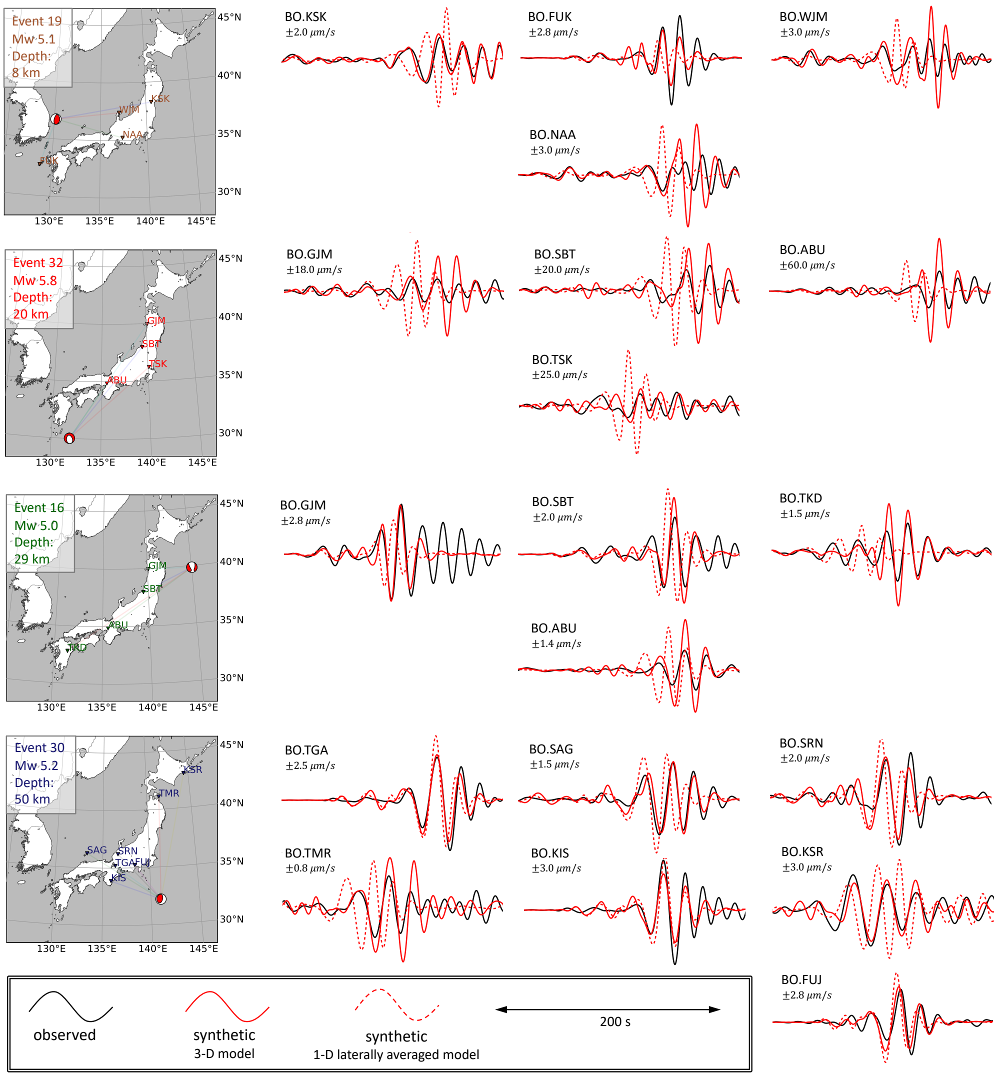

Figure 2. Representative collection of observed waveforms (black), synthetic waveforms computed for the final 3-D model (solid red) and synthetic waveforms computed for the 1-D laterally averaged model (dashed red). The waveforms are filtered between 15 – 80 s. We show the vertical component of the waveforms for four events and selected stations, with the source-receiver configuration specified in a separate map for each earthquake. Event information, shown in the top left corner of each map, are NIED CMT solutions (Fukuyama et al., 2001).

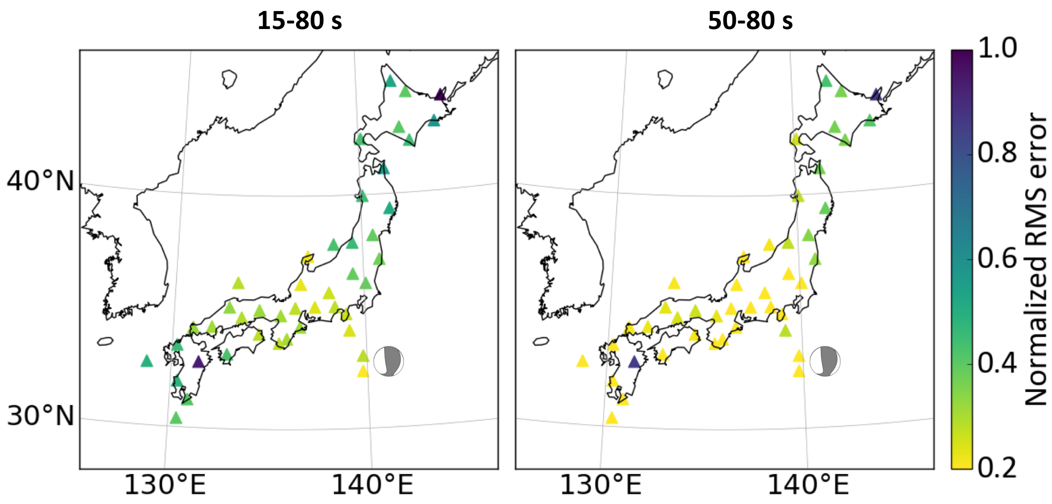

Figure 3. Root-mean square (RMS) error between the vertical component of the observed data and the synthetic seismograms calculated for the Global Centroid-Moment-Tensor (GCMT) solution for the M_w 5.2 event at 50 km depth depicted as a grey focal mechanism. The waveforms are filtered between 15–80 s (left) and 50–80 s (right). Misfits are normalized to the largest value of both scenarios. While longer-period data are well explained at the majority of the stations, more variation in misfit is present at shorter periods. Generally, stations in central Japan exhibit a very good fit, while those further away from the event, such as in Hokkaido or Kyushu, are characterized by slightly elevated misfits.

3 Forward problem

3.1 Representation theorem

According to the representation theorem, the i-component of the displacement field $\boldsymbol{u}$ in a point-source configuration can be expressed as a convolution of a time-dependent moment tensor $M_{nq}(t)$ and Green's strains $G_{in,q}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau)$ (Aki & Richards, 2002):

$$u_i(\boldsymbol{x}, t) = M_{nq}(t) * G_{in,q}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau), \tag{2}$$

where Green's function $G_{in}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau)$ is the i-component of the displacement field recorded at location $\boldsymbol{x}$ and time t due to an impulse in n-direction at location $\boldsymbol{\xi}$ and time τ, and $G_{in,q}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau)$ is a spatial gradient of Green's function with respect to the q-coordinate of the source location, with $,q$ denoting a derivative with respect to ξ_q. Einstein notation is implied.

In practice, it is often assumed that separate moment tensor components have the same time dependence, or the same source time function $s(t)$ (e.g., Dziewoński et al., 1981; Ekström et al., 2012; Mustać & Tkalčić, 2016; Takemura et al., 2020; Zhu & Zhou, 2016), in which case a component of the displacement field can be expressed as:

$$u_i(\boldsymbol{x}, t) = M_{nq}\, s(t) * G_{in,q}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau). \tag{3}$$

The source duration of $\sim$ M_w 5 events, which we consider in this study, is usually a few seconds long (e.g., Vallée & Douet, 2016). Hence, we assume an instantaneous

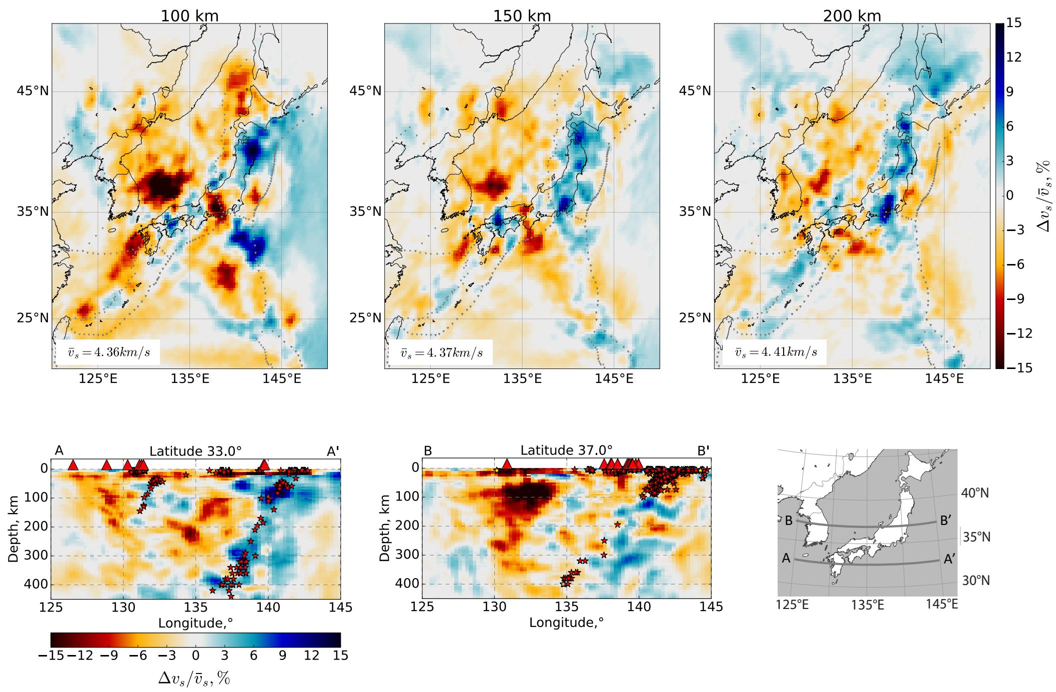

Figure 4. Percentage perturbations of the isotropic S velocity, computed as $\frac{v_s - \bar{v}_s}{\bar{v}_s} \times 100$ %, where $\bar{v}_s$ is the lateral average of v_s for each depth. **Top panels**: horizontal slices, with dashed grey lines representing plate boundaries. **Bottom panel:** vertical cross-sections. Red and yellow stars represent earthquakes since 1997 and earthquakes used in this study, respectively, within 1° of the slice. Red triangles represent Holocene volcanoes (Siebert et al., 2010).

source time function, which is a sufficient approximation for the shortest periods we work with, i.e., 15 s, which was also shown in the pilot study by Fichtner and Simutė (2018). Assuming the same source time function for all events and all moment tensor components, we can convolve $s(t)$ with the Green's strains at the time of computation, in which case the displacement field becomes a linear combination of convolved Green's strains scaled by the moment tensor elements.

To ensure a rapid forward problem for probabilistic inference, we pre-compute and store the Green's strains in a database, taking advantage of spatial reciprocity. The merits of reciprocity for the computations of Green's functions have also been exploited in previous studies (e.g., Eisner & Clayton, 2001; Zhao et al., 2006; Lee et al., 2011; Hejrani et al., 2017; Okamoto et al., 2018; Takemura et al., 2020). The reciprocal formulation of Eq. 3 can be found in Fichtner and Simutė (2018).

3.2 Database of Green's strains

We compute Green's strains numerically with the spectral-element solver SES3D (Fichtner et al., 2009b; Gokhberg & Fichtner, 2016). Enabled by reciprocity, we treat seismic stations as virtual sources and save the wavefield across the actual source area of interest, i.e., the Izu-Bonin trench. To ensure a continuous representation of the wavefield within the domain, we store the wavefield on Gauss-Lobatto-Legendre (GLL) points of the fourth-order spectral-element method (SEM) grid and use the built-in polynomial interpolation of SEM to extract the wavefield for any spatial coordinate. This contrasts with the common practice of storing the Green's functions on a pre-defined grid (e.g., Hejrani et al., 2017; Lee et al., 2011; Vackár et al., 2017; Takemura et al., 2020), where one has to implement an interpolation routine or deal with a finite number of discrete locations and possibly limit the spatial resolution of the earthquake location. Storing the wavefield itself allows us to extract the strains for any potential source location and be exempt from any additional parametrisation effects. We compute the database for over 50 selected F-net broadband stations (Fig. 5) uniformly distributed across the network (National Research Institute for Earth Science and Disaster Resilience, 2021).

The source area of interest extends between 140°−143° E, 30°−35° N, and down to 110 km depth. The downsampled wavefield with a time increment of 2 s takes 27 Gb of space for a single virtual source, and the total storage requirements are 4.2 Tb. The database is stored on Piz Daint supercomputer in the Swiss National Supercomputing Centre, which we use to rapidly perform the inversions (Swiss National Supercomputing Center, 2021).

4 Earthquake selection

4.1 Earthquakes in the area

The Izu-Bonin trench marks the boundary between the subducting Pacific plate and the Philippine Sea plate. The trench is situated nearly linearly from north to south. It is a steeply dipping subduction zone, with the angles of the Wadati-Benioff zone between 50° and 70° (Faccenna et al., 2018). Along the Izu-Bonin slab, seismicity extends from the shallow surface down to the transition zone in the south and $\sim$ 410 km depth in the north (Dziewoński et al., 1981; Ekström et al., 2012; Seno & Eguchi, 1983; Hayes et al., 2012; Hayes, 2018). Following the global trend, the majority of events are located in the upper $\sim$ 60 km (Kong et al., 2018; Hasegawa, 2011) (Fig. 5). At this depth, seismicity primarily occurs as a low-angle interplate thrust faulting, reflecting the relative motion of the convergent plates (Hasegawa, 1990, 2011). Deeper down, earthquakes mostly take place within the slab (Hasegawa, 1990, 2011). In the Izu-Bonin arc – an old plate subduction zone – these intraslab events have the compressional axis predominantly oriented in the dip direction (Hasegawa, 2011). In the overriding plate, the compression is

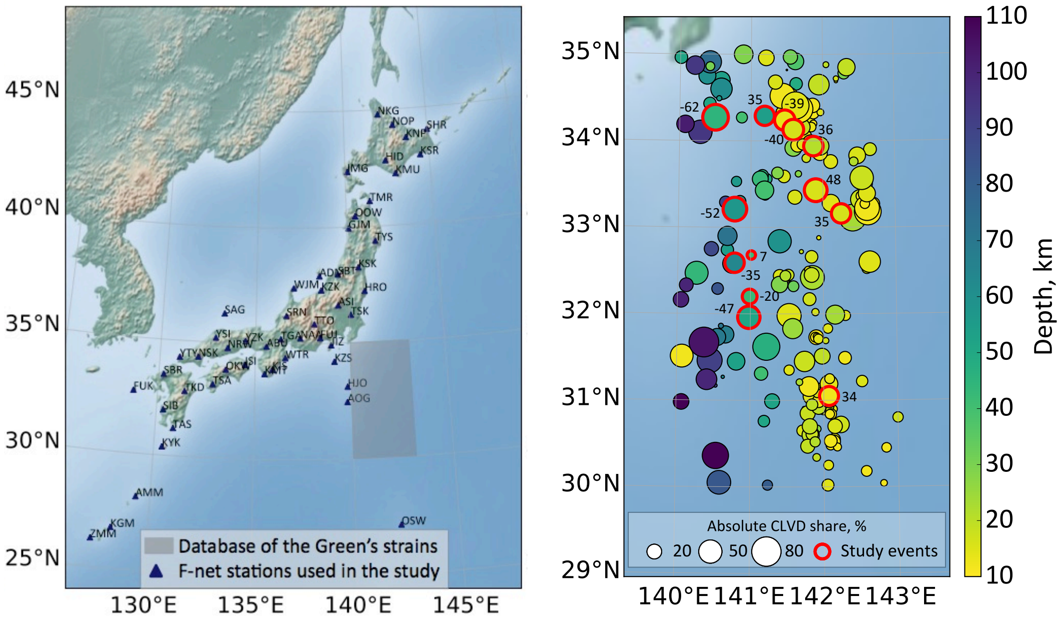

Figure 5. **Left:** setup of the stations and the source area used in the source inversion. Receiver-side Green's strains were computed from each seismic station, acting as a virtual source, and stored within the shaded source area. The strain database extends from the surface to 110 km depth. **Right:** distribution of earthquakes within the horizontal extent of the source area as given in the GCMT catalog between 1997 and 2020 (The Global CMT Project, 2021). The color of the circles corresponds to the depth of the earthquake, and the size to the absolute share of the CLVD component. Events used in this study are outlined in red. Depth scale saturates at a maximum depth.

accommodated by intraplate thrust fault or strike-slip fault earthquakes, with compressional axis oriented in the direction of plate convergence (Hasegawa, 1990).

Given the complex nature of the subduction zone, earthquake mechanisms are diverse. Notably, there are numerous strongly non-double-couple (non-DC) events, with compensated linear vector dipole (CLVD) component reaching up to 80 % of the total moment (Fig. 5). There are physical explanations for CLVD mechanisms, such as simultaneous faulting of two non-parallel planes (Kuge & Kawakatsu, 1993), or complex faulting with any deviation from unidirectionality in terms of a strike, dip or rake, with volcanic caldera collapse being a perfect example how many nearly simultaneous slips on a curved fault result in an effective vertical-CLVD earthquake (Shuler et al., 2013; Nettles & Ekström, 1998; Fichtner & Tkalcic, 2010). However, very often an apparent CLVD component is an artifact caused by a modeling error. Incorrect Earth structure, especially around the hypocenter of the earthquake, has a significant influence on moment tensor estimation (Shuler et al., 2013; Burgos et al., 2016). For an intuitive understanding, one can think in terms of a first-polarity inversion and a simple double-couple earthquake. The take-off angle depends on the velocity structure in which the earthquake is embedded. When the take-off angle is incorrect, the inferred pressure (P) and tension (T) axes, which correspond to the middle of dilatational and compresional quadrants, respectively, are also incorrect (e.g., Newrkla et al., 2019). The inconsistencies can go a long way, such that the mechanism can no longer be explained by a double-couple, and the errors in modeling are then compensated by introducing a CLVD component. Inability to clearly distinguish between the physical versus apparent CLVD component inhibits our understanding of earthquake physics, while possibly incorrect focal mechanisms hinders the accurate delineation of the local tectonic setting. Hence, by incorporating complex Earth structure in our study we expect to see whether the CLVD component is a physical feature of the earthquake or an artifact due to modeling errors.

4.2 Study events

We study events of moderate magnitudes, which fall within the area of our strain database (Fig. 5) and have a significant CLVD component. Preference is given to more recent earthquakes away from the database boundaries. We selected 13 events from the Global Centroid-Moment-Tensor (GCMT) catalog (Dziewoński et al., 1981; Ekström et al., 2012; The Global CMT Project, 2021). The earthquakes have moment magnitudes between M_w 4.8–5.3, are distributed within 13 km and 64 km depth with CLVD component ranging between 7 % and 62 % (in absolute sense), with a median value of 36 % (Fig. 5, 6, Table S1). The moment tensor decomposition is done after Vavryčuk (2015), which has been adopted in previous studies (e.g., Vackár et al., 2017; Yu et al., 2019; Shang & Tkalčić, 2020). We note that the deviatoric part of the moment tensor can be decomposed into a number of equivalent force combinations, and hence the choice is subjective (e.g., Jost & Herrmann, 1989).

5 Inverse problem

5.1 Bayesian inference

We work in the Bayesian framework, where according to Bayes' theorem (Bayes & Price, 1763) the posterior probability density $\pi(\boldsymbol{q}|\boldsymbol{d})$ of the model vector $\boldsymbol{q}$ given the data $\boldsymbol{d}$ is:

$$\pi(\boldsymbol{q}|\boldsymbol{d}) = k\pi(\boldsymbol{d}|\boldsymbol{q})\pi(\boldsymbol{q}). \tag{4}$$

Bayes' theorem provides a framework to enhance the existing knowledge, or the prior probability density $\pi(\boldsymbol{q})$, with the new information from the data, that is, the likelihood $\pi(\boldsymbol{d}|\boldsymbol{q})$ (Mosegaard & Tarantola, 2002; Mosegaard & Sambridge, 2002; Sambridge & Gal-

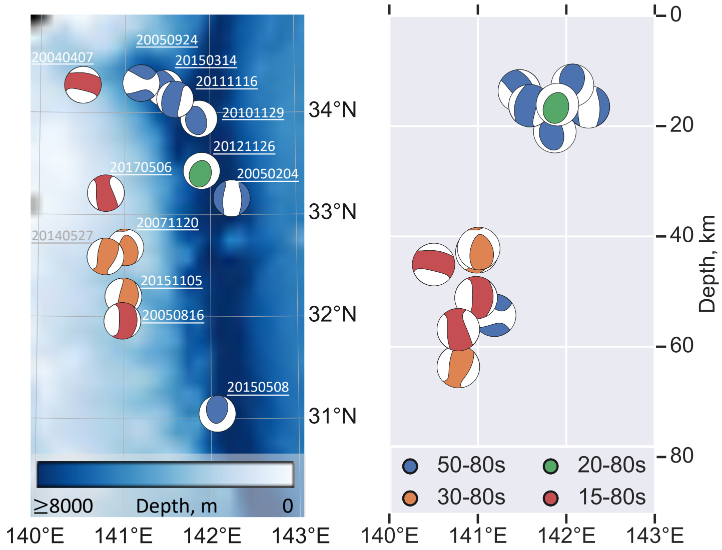

Figure 6. Distribution of earthquakes selected for this study. Earthquakes are plotted in terms of their focal mechanisms, with colors representing the shortest acceptable periods used for the inversion (see section 5.4 for more detail). **Left:** horizontal distribution of the events and their IDs plotted on the bathymetric map. **Right:** depth distribution of the study events. Deeper events can be modeled over a wider frequency range compared to the shallow ones.

lagher, 2011; Fichtner, 2021). The likelihood term contains information on the data fit, i.e., how well the current model can explain the data. A constant k ensures the integral of the posterior probability density over the model space is equal to 1 (e.g., Sambridge & Mosegaard, 2002; Mustać & Tkalčić, 2016; Staehler & Sigloch, 2014).

We express the likelihood $\pi(\boldsymbol{d}|\boldsymbol{q})$ as the exponential function of the negative L_2 misfit between the synthetic and the observed waveforms, $\mathbf{s}$ and $\mathbf{d}$, respectively:

$$\pi(\boldsymbol{d}|\boldsymbol{q}) \propto \exp\left(-\chi\right), \tag{5}$$

$$\chi = \frac{1}{2}(\mathbf{s}(\mathbf{q}) - \mathbf{d})^T \boldsymbol{C_D}^{-1}(\mathbf{s}(\mathbf{q}) - \mathbf{d}), \tag{6}$$

where $\boldsymbol{C_D}$ denotes the data covariance matrix.

5.2 Hamiltonian Monte Carlo

We pursue Bayesian inversion with the Hamiltonian Monte Carlo method. Originally introduced as *hybrid* Monte Carlo (Duane et al., 1987), the method derives from molecular dynamics simulation, used to study the properties of many-body systems by solving Newton's equations of motion (e.g. Xu & Li, 2008; Alder & Wainwright, 1959; Neal, 1993). In short, HMC can be regarded as an efficient proposal mechanism, which relies on exploiting gradient information of the model parameters. The main idea of the algorithm is to follow a contour of high-probability, which is achieved by balancing the gradient, or the *force*, by an artificially introduced *momentum*.

To set the stage for HMC, we first expand our model parameter space, described by the N_q-dimensional position vector $\boldsymbol{q}$, with auxiliary momentum parameters $\boldsymbol{p}$. For physical intuition of the sampling process one could imagine a mechanical particle in phase space. To propose a new sample, a particle is set into motion by randomly assigning momentum to each model parameter. The particle then travels along the trajectory for some artificial time τ. The end of the trajectory serves as a new proposal. By marginalizing over the artificially introduced momentum, we can retain only the position variables, i.e., the physical part of the phase space, which we are actually interested in. Mathematically, the trajectory is governed by Hamiltonian dynamics:

$$\frac{dq_i}{d\tau} = \frac{\partial K}{\partial p_i}, \qquad \frac{dp_i}{d\tau} = \frac{\partial U}{\partial q_i}, \qquad i = 1, ..., N_q, \tag{7}$$

where potential energy U is expressed as:

$$U(\boldsymbol{q}) = -\ln \pi(\boldsymbol{q}|\boldsymbol{d}), \tag{8}$$

and kinetic energy must be defined by the implementation. In this study we use:

$$K(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T \mathbf{M}^{-1} \boldsymbol{p}, \tag{9}$$

where the positive-definite mass matrix $\mathbf{M}$ is a tuning parameter, which, generally speaking, acts as a scaling parameter to ensure that momentum is tailored to the sensitivity of each model parameter. This in turn allows us to explore the space equally well for each parameter. The solution of Hamilton's equations throughout the artificial time τ represents the evolution of the model in phase space. The discretization of artificial time and the total length of the trajectory, L, are the tuning parameters of HMC. The dynamics conserves the total energy H throughout the Hamiltonian trajectory:

$$H(\boldsymbol{q}, \boldsymbol{p}) = U(\boldsymbol{q}) + K(\boldsymbol{p}). \tag{10}$$

To solve the differential equations (Eq. 7), we numerically integrate using a leapfrog algorithm, which preserves the volumes of regions of phase space, meaning, that by moving from one region in phase space to another, the points retain the same volume and reversability in time, two properties important for HMC (Neal, 1993, 2011). The total energy, on the other hand, is not conserved by the leapfrog alogrithm, and this affects the acceptance rate of the proposed samples.

The algorithm is performed in steps, starting with some model $\mathbf{q}$:

1. Draw momentum values from the multivariate normal distribution $\exp\left(-\frac{1}{2}\boldsymbol{p}^T \mathbf{M}^{-1} \boldsymbol{p}\right)$.
2. With $\mathbf{q}$ and $\mathbf{p}$ specified, solve Hamilton's equations (Eq. 7).
3. The end of the trajectory marks a newly proposed sample in terms of $\mathbf{q}(L)$ and $\mathbf{p}(L)$. After evaluating the total energy of the new sample, H, the model is accepted with probability:

$$\Pi_{accept} = \min\left[1, \frac{\exp[-H(\boldsymbol{p}(L), \boldsymbol{q}(L))]}{\exp[-H(\boldsymbol{p}, \boldsymbol{q})]}\right]. \tag{11}$$

4. Repeat the procedure from step (1). If the sample is accepted, use $\mathbf{q}(L)$ as a new starting point, otherwise, return to the beginning of the trajectory and reuse model $\mathbf{q}$.

For technical aspects and choice of tuning parameters of HMC we refer the reader to Fichtner and Simutė (2018).

5.3 Inversion parameters

In this study we seek a centroid-moment tensor solution, which means we simultaneously infer a centroid location, centroid time and a moment tensor of an earthquake (Dziewoński & Woodhouse, 1983). We invert for a full moment tensor, i.e., six independent components M_{ij}. By not imposing any constraints on the faulting mechanism, we allow the mechanism to be determined freely by the data. If a parameter or a combination of parameters cannot be constrained by the data, i.e., it lies in the null space of the model space, this shall be seen in the uncertainties provided by the probabilistic inference. For the comparison purposes, we also run a separate inversion for each case imposing a zero-trace constraint on the moment tensor, which denotes a source with no volumetric component.

Working with the moment tensor components M_{ij} is a subjective choice, and various alternatives exist (Tape & Tape, 2013). An advantage of the probabilistic approach used in this work is that the subjective component is explicit, and that it can be modified via a simple re-parameterisation of the involved probability densities. Hence, if needed, results can easily be presented in any different parameterisation, without suffering from subjective regularisation bias.

For the misfit computation we select measurement windows from the full waveforms manually and prescribe weights in order to preserve information carried by the body waves, which would otherwise be suppressed by the larger-amplitude surface waves. Assuming uncorrelated Gaussian data noise, our data covariance matrix $\boldsymbol{C_D}$ is a diagonal matrix, entries of which we conservatively estimate from the pre-signal noise. We express prior probability density on model parameters as Gaussian distributions with standard deviations around ten times larger than the parameter mean for the moment tensor elements ($1{\times}10^{17}$ N m), and 2°, 20 km, and 2 s for horizontal location, depth, and centroid time, respectively. The priors are intentionally made very wide in order to put emphasis on the constraints provided by the data and to not bias the inference too much by the prior. As a prior mean we commonly use the solution provided by the GCMT (Dziewoński et al., 1981; Ekström et al., 2012). However, for some events (IDs 20050816, 20040407) we update the location first by running a preliminary inversion with a reduced number of samples, and then use the maximum-likelihood model from this inversion as a prior mean for the main inversion. We refer to the maximum-likelihood model, as the one having the minimum potential energy U (Eq. 8).

5.4 Multi-period band inversion

Seismic waves traveling through complex geology for many wavelengths accumulate complicated path effects, such as frequency-dependent scattering or focussing in the presence of seismic velocity heterogeneities, which become more pronounced at shorter periods (e.g Igel & Gudmundsson, 1997; Igel et al., 2002; Ferreira & Woodhouse, 2007). While our tomographic model can explain the majority of the waveforms in the 15–80 s period band, some complexities remain unaccounted for. This is primarily because strongly heterogeneous geology (e.g., accretionary prisms in the subduction zones) cannot be fully resolved by our limited data, especially due to non-uniform source-receiver distribution.

As expected, our model explains longer-period data better (Fig. 3), but omitting shorter periods reduces the information content carried by the waveforms. Hence, we are faced with a trade-off between a very good long-wavelength Earth model and the available short-period information, which is necessary to constrain a full moment tensor, including its volumetric component.

In our approach we perform multi-period band inversions, which means that we invert the same event using different period data, i.e., 15–80 s, 20–80 s, 30–80 s, and 50–80 s. In the analysis we only consider those event – period-band configurations for which

Figure 7. Left: L_2 misfit (Eq. 6) between the observed data and the synthetics for the maximum-likelihood model from the inversion at each of the four period bands. The misfits are computed for the vertical component and are normalized to the largest value for visualization. Misfits have a high variance at short periods, but converge to a similar value for long-period inversions, indicating that data are not equally well explained for events at short periods. Right: waveform fit for two selected earthquakes in different period bands. The acceptable inversion period band of event 20050816 (in blue) is 15–80 s. It has a good waveform fit at short, as well as long periods. The acceptable inversion period band of event 20150314 (in green) is 50–80 s. For all events, waveforms are better explained with increasing periods.

an adequate waveform fit between the synthetic seismograms for the maximum-likelihood model and the observed data is achieved. The adequate waveform fit is guided by the L_2 misfit between the observed data and the synthetic data (Eq. 6). The procedure is not purely automatic and requires human interaction. The uncertainty of the L_2 misfit comes through its dependence on the estimated data noise and its assumed distribution. While this is fine in the inversion, where each event is considered separately, the absolute comparison across events might be problematic, because it depends on the subjectively assigned data error. That is why in addition to the L_2 misfit, we also evaluate traces visually.

At shorter periods, the misfits vary more significantly than at longer periods (Fig. 7). Waveforms of some events, which tend to be deeper, are explained better than those from other events, which tend to be shallower. This is illustrated in the right column of Fig. 7, where the waveforms for event 20150314 (GMCT depth 13.6 km) are not well explained at short periods, but the fit becomes adequate at long periods. Event 20050816 (GMCT depth 51.1 km), on the other hand, has an adequate waveform fit throughout all the period bands. In our approach by varying the period, we seek an Earth model, which could largely explain the observed data and could therefore, be used for the source inversion.

In the following section we will investigate the results of three events inverted with 15–80 s, 30–80 s, and 50–80 s data and discuss the differences in solutions as seen by different period data. We will then group the events according to their shortest acceptable inversion period and provide a general overview of events from each period band: three aforementioned events inverted with 15–80 s period data, one event with 20–80 s, three events with 30–80 s, and six events inverted with 50–80 s period data.

Figure 8. Ensembles of the focal mechanisms for three events inverted with 15{80 s, 30{80 s, and 50{80 s period data. Gray lines within the beachballs represent every 100th accepted model and the colored mechanisms correspond to the maximum-likelihood model. Red beachballs represent unconstrained inversions, and the blue ones inversions where the isotropic component is xed to zero. Gray beachballs at the top represent the GCMT solution. The double-couple component of the constrained inversion exceeds that of the GCMT for all events in all period bands. Generally, the DC component reaches the highest value for the shortest-period data inversion and decreases with increasing period band.

Figure 11. Comparison of selected trade-offs and marginal probability density functions for 15–80 s (top) and 50–80 s (bottom) period data inversion of event 20040407. The limits for each parameter depend on the corresponding standard deviations, and are set from $\mu - 3\sigma$ to $\mu + 3\sigma$, where μ is the mean and σ is the standard deviation of the distributions. When shorter-period data are used, the posteriors are more complicated (e.g., M_{yz}; M_{xx}) and sometimes have several maxima. Because the multi-parameter distributions are not exactly Gaussian, maximum-likelihood models do not correspond to the mean values and for some parameters lie outside of the main probability densities. Depth trades off with the majority of the moment tensor components. For most of the parameters the trade-offs remain consistent for short- and long-period data inversions, however it is not always true (e.g., M_{yz} and M_{zz}). Although for this particular event, trade-offs seem to be stronger when short-period data are used, it is not a general feature across other events.

6.2 Waveform fit

In Fig. 13, we present the waveform fit for two events inferred with 15–80 s period data. Compared to the GCMT solution, the fit improves by 4 % for event 20050816 and 7 % for event 20040407 for the unconstrained inversion. The maximum-likelihood solution from the constrained inversion usually gives slightly worse fit compared to the unconstrained one, the difference is 1 % for the event 20040407 presented in the lower panel of Fig. 13. Although the numeric waveform fit improvement is relatively small, its effect on the source mechanism is significant. An improvement of 4 % for event 20050816 means 21 % increase in DC component (Fig. 8) and a depth relocation of 7 km (Fig. 9). Hence, large variations in source parameters are hidden in the subtle waveform differences, which are possible to extract only by virtue of a good Earth model.

6.3 Statistical analysis

Here we present all events inverted in their shortest acceptable period band (see section section 5.4) and discuss those features of the results, which manifest across all the inversions.

mean vector $\boldsymbol{\mu}$ is multiplied by an $N\times1$ vector of ones, and $\mathcal{E}$ denotes the expected value of the product in the square brackets. Because our model parameters are incommensurable, i.e., they have diverse physical units, such as s or Nm, and variances with orders of magnitude ranging from 1×10^{-4} to 1×10^{30}, we work with a correlation matrix (Chave, 2017). The correlation matrix is characterized by the normalized covariances, but nevertheless, retains the inter-parameter trade-offs:

$$corr(\boldsymbol{X}) = \begin{pmatrix} 1 & \rho_{12} & . & \rho_{1K} \\ \rho_{21} & 1 & . & \rho_{2K} \\ . & . & . & . \\ \rho_{K1} & \rho_{K2} & . & 1 \end{pmatrix} \tag{13}$$

where $\rho_{jk} = \sigma_{jk}/\sqrt{\sigma_{jj}\sigma_{kk}}$.

The first principal component, or an eigenvector with the corresponding largest eigenvalue, is responsible for the largest variance of the distribution. In our case this corresponds to the greatest uncertainty axis. The best-constrained direction of the multi-dimensional distribution is represented by the smallest PC, indicating the narrowest extent of the distribution. The exact share of the explained variance is proportional to the size of the eigenvalue. If there are no dominant eigenvalues and all of them are of a comparable size, it means that all the principal components, or the effective parameters, are resolved with a similar (un)certainty. The correlation matrix is then close to being proportional to the identity matrix, resembling a multi-dimensional sphere (Chave, 2017). In a dataset of observations, if some threshold variance, typically set between 70 % and 90 %, can be explained by much fewer PCs than a number of original parameters, one could retain only the most important principal components, and benefit from the reduced dimensionality of the problem (Chave, 2017).

We study principal components for the unconstrained inversion of each study event at its corresponding shortest acceptable period band as presented in section 6.3.1. The left column of Fig. 17 illustrates the cumulative explained variance by the principal components for each event. We see that events inverted at longer periods (30–80 s, 50–80 s) reach the 70 % variance threshold with the first three and 90 % with the first five principal components, while those inverted at shorter periods (15–80 s, 20–80 s) need four for 70 % and seven for 90 % variance threshold. In other words, the eigenvalues from the long-period inversions show larger variability than those from the shorter-period inversion. From this we can deduce that the shape of multi-dimensional distribution for longer-period inversions diverge from the sphere more than for the short-period inversions. This means that for an event inverted at long periods, some axes (e.g., PC 0) are much more difficult to constrain than others (e.g., PC 9), while at shorter periods, they can all be constrained more equally.

In order to ascertain whether these certainty and uncertainty axes are similar across all the inversions, we investigate similarity of the principal components among the different events by computing the dot product between PCs associated i) with uncertainty (PC 0–PC 4) and ii) certainty (PC 5–PC 9) within the same period band inversions. Putting a threshold of 30°, we find no consistency among the axes corresponding to either the smallest or the largest eigenvalues. In other words, principal components are pointing to different directions for each event and no linear combination of physical parameters can be generalized to be the least- or the best-constrained direction in the model space. This suggests that we cannot easily reduce the dimension of the problem by ignoring some parameters, because the eigenvectors are different for each event. Because the data and the azimuthal coverage are similar for all the study events, it is likely that the source mechanism or the centroid location are responsible for the fact, that each event is characterized by a very different set of principal components.

7 Discussion

7.1 Increase of double-couple component

Accounting for heterogeneous Earth structure enables us to use shorter-period data (15–80 s, 20–80 s) in regional source inversion. This leads to a higher double-couple component compared to GCMT solutions, and it suggests that at least part of the CLVD component, given by the GCMT catalog, is likely to be an artefact caused by forward modeling errors. A similar trend of increasing DC component when 3-D Earth structure is taken into account has been shown by Hejrani et al. (2017) and X. Wang and Zhan (2019) for earthquakes in the Papua New Guinea and Los Angeles regions, respectively.

7.2 Power of subtle waveform differences

Significant changes in source parameters, both in terms of moment tensor and location, may occur despite relatively small waveform differences. In one of the examples we showed that an overall waveform fit improvement of 4 % led to 21 % increase in DC component together with a depth relocation of 7 km. This implies that some source parameters or combinations thereof are highly sensitive to subtle waveform differences and Earth structure, and can therefore only be resolved with a reliable 3-D Earth model.

7.3 Effective source solutions

In this study we have presented effective, i.e., period-dependent, point-source solutions, inferred with different-period data. In the following, we discuss means of how varying frequency data can lead to effective source parameters.

7.3.1 Source complexity

Spatial and temporal complexity of the source have been previously suggested to explain the discrepancy between short- and long-period source mechanisms (e.g., Wallace et al., 1982; Grandin et al., 2015; Frankel, 2013). For example, it has been argued that asperities on the fault surface generate shorter-period seismic waves, while the overall faulting episode is represented by the long-period waveforms (Wallace et al., 1982). However, a typical M_w 5 event, such as those selected for our study, should not exceed several seconds in half-duration and a few kilometers fault surface radius, which makes it unlikely for 15 s period data to constrain a subfaulting episode, as the periods are much longer than the expected duration of the earthquake (Ekström et al., 2012; Hanks, 1977; Eshelby & Peierls, 1957).

7.3.2 Near-source anisotropy

Another possibility, which could explain effective source solutions, is related to the near-source heterogeneities and specifically, anisotropy. Anisotropy arises from different, scale-dependent mechanisms (e.g., Backus, 1962; Kawakatsu et al., 2009; N. Wang et al., 2013). Such a frequency-dependent nature of observed anisotropy has indeed been reported beneath Japan (Wirth & Long, 2010) and other subduction zone regions (e.g., Fouch & Fischer, 1998; Greve & Savage, 2009). Anisotropy in the immediate vicinity of the source can affect wave propagation in such a way that a purely isotropic event might appear to have excited shear waves, while a shear earthquake might appear to have had a non-DC component (e.g., Vavryčuk, 2004; Li et al., 2018). The effective source solutions might arise from the fact that the fine-scale anisotropy is only captured by the shorter-period waves, while longer-period waves sample an effective medium over larger scales. Therefore, at short-period inversions (15–80 s), where the fine-scale anisotropy around the source is accounted for, the apparent CLVD component decreases, while at long periods (50–80 s), it remains relatively high.

7.3.3 Information content

Long-period data may not contain enough information to properly constrain the source parameters of relatively small-magnitude earthquakes. Firstly, data are more correlated at longer-periods and hence, carry less independent information. This results in inherent trade-offs between moment tensor components, such as M_{zz} and $(M_{xx}+M_{yy})$, which trade off in the case of long-period surface wave inversion (Dufumier & Rivera, 1997; Fitch et al., 1981). Secondly, the amplitude spectrum of small-magnitude events tends to decrease with increasing period (Aki, 1967), and hence, the signal-to-noise ratio at long periods may therefore, be diminished.

Long-period data used in this study, i.e., body and surface waves in the 50–80 s period band, are comparable to the data used to construct the GCMT catalog. The reported minimum periods of body and surface waves, used to constrain the earthquakes chosen for our study, are 40 s and 50 s, respectively (The Global CMT Project, 2021). Although the azimuthal coverage in the GCMT inversion may be more complete than in our study, due to the available teleseismic data, the lack of seismic receivers in the north-western Pacific Ocean is evident (Ekström et al., 2012). Similarity between data periods could potentially explain why our long-period source solutions are in a close agreement with the GCMT predictions.

7.4 Effect of the prior

The source parameters constrained with long-period data are often very close to the prior model, i.e., the GCMT solution, which we use for the majority of the inversions. In the Bayesian framework, the fact that the posterior closely resembles the prior can mean two things: either the likelihood is very similar to the prior and hence prefers the same solution, or the data constraints are weak, and the posterior is dominated by the prior probability density distribution.

Although stochastic algorithms are direct search methods, theoretically giving us a chance to obtain probability density functions of a full model space, its practical implementation might be very expensive, especially in a high-dimensional space. With the aim to speed up the convergence and to alleviate computational costs, we introduced a modified version of Hamiltonian Monte Carlo. If well tuned, HMC can be a very efficient sampling algorithm for its gradient-based approach of proposing the samples. The derivative should in principle be computed at every point along the Hamiltonian trajectory. However, this is expensive, and in our formulation we suggest to approximate the derivative around a prior mean in model space, instead (Fichtner & Simutė, 2018). The derivative is exact for the parameters linearly related to data (i.e., the moment tensor components), but it is an approximation for nonlinear parameters (i.e., centroid). We note that this approximation only concerns the samples drawn to be proposed, and not the acceptance criterion, which should still ensure that even with this approximation, only relevant models are accepted. Furthermore, if we let our sampler run for an infinite amount of time, the way samples are proposed would not matter. However, during the finite run time, the posterior might actually be biased towards the point in model space around which the approximation is performed, which in our case is often prior mean.

7.5 Data and modeling errors

In this study we pragmatically treated observed data noise and forward modeling errors. We estimated data noise from the pre-signal recordings, and assumed it to be normally distributed without spatial or temporal correlation. For this conceptual demonstration we also did not account for the forward modeling errors. Efforts to properly incorporate modeling errors and data noise covariances in stochastic inversions were made by e.g., Staehler and Sigloch (2014, 2017); Vackár et al. (2017); Hallo and Gallovič (2016); Duputel et al. (2012).

7.6 Trade-offs between structure and source

Because of the trade-offs between Earth structure and earthquake source parameters, it is challenging to constrain them independently (e.g., Morales-Yáñez et al., 2020; Hjörleifsdóttir & Ekström, 2010; Hejrani et al., 2017). In this study, we proceed with a two-step approach: first we constrain the structure, then, with a suitable Earth model at hand, we aim to recover improved source parameters. In an idealized scenario, where the probability density distribution between the source and the structure parameters would be Gaussian, this procedure would look rather straightforward (Fig. 18). However, the real world presents many complications. Firstly, we do not have much constraints on the model space of the Earth structure except for the least-squares solution. Secondly, such a unimodal distribution might be an oversimplification. Lastly, the probability density distribution is likely to be frequency-dependent, with increasing complexity at increasing frequencies. Therefore, locating a global minimum in the source – structure space might actually be a very difficult task, especially at shorter periods.

8 Conclusions

We presented results of a probabilistic seismic source inversion for 13 small-to-moderate magnitude offshore earthquakes at the Izu-Bonin trench. The inversions were conducted using fully heterogeneous, radially anisotropic Green's functions and the Hamiltonian Monte Carlo sampling algorithm. We simultaneously inferred centroid location, centroid time and six independent moment tensor components, and ran a separate inversion constraining the volumetric component to zero for comparison purposes. With the goal to use a sufficiently good Earth model, we varied the minimum inversion period and limited ourselves to those period bands, for which the data fit between the observed data and the maximum-likelihood solution was sufficiently good.

Accounting for 3-D Earth structure at short periods (15–80 s, 20–80 s) generally leads to an increase in DC component compared to the GCMT solution (Fig. 14). This suggests that at least some part of the non-DC component in the GCMT catalog might be apparent – resulting from unmodeled Earth structure. Events inverted at longer periods (e.g., 50–80 s) (Fig. 15), do not show a significant change in mechanism or centroid location and stay close to the GCMT predictions. We have presented several possible mechanisms to explain the effective solutions, the most likely being near-source propagation effects, which cannot be resolved by long-period data, or weakening data constraints with increasing periods.

Constraining isotropic component to zero is a common practice in order to stabilize the inversion and to prevent unexplained waveform differences from mapping into the additional parameter. Here, we observe that owing to the 3-D Earth model, isotropic component remains weak even when unconstrained.

The diagonal components of the moment tensor, which are responsible for the volumetric change, are significantly less constrained, compared to the off-diagonal ones (Figs. 10, S8). From the events which could be inverted in all period bands, we also see that shorter-period data constrain all the source parameters better than the long-period data (Figs. 10, S8). Posterior probability density distributions of shorter-period data inversions appear multimodal and are more complicated than those of long-period. This illustrates non-uniqueness of short-period source inversions and highlights the need for stochastic approaches.

We note that finding a solution, which has a significantly better waveform fit than that provided by the GCMT catalog, is a challenging task. In our examples, the waveform fit only improves by several percent (Fig. 13). However, a small change in the waveforms brings about a significant change in the source solution. In other words, large variations in source parameters are hidden in the subtle waveform differences, which are possible to extract only by virtue of a good Earth model.

To aid the study of the multi-dimensional posterior, we perform principal component analysis. We include all the study events at their shortest acceptable inversion period band, with the aim to retrieve the best- or the least-constrained direction in model space. By comparing the principal components (eigenvectors) corresponding to either the smallest or the largest eigenvalues, we find no consistency among different events from the same period-band inversion group. Each event is characterized by a very different set of principal components, and no linear combination of physical parameters stands out as the least- or the best-constrained direction in model space.

In this conceptual study, we detailed the methodology for a probabilistic source inversion using 3-D Green's functions and presented a proof-of-concept catalog of source solutions. Such an approach allows us to better constrain source characteristics and comes with the ensemble statistics, such as uncertainty limits and inter-parameter trade-offs. Inferred source parameters contribute to our understanding of the regional seismotectonics and earthquake physics and can also be fed back into and, potentially, improve tomographic studies. The method could be run in production mode for any part of the world, for which a reliable 3-D Earth model is available.

9 Open Research

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 Northeast China Extended Seismic Array, the Global Seismograph Network, and the Korean Seismic Network, made available by the IRIS Data Management Center (http://ds.iris.edu/ds/nodes/dmc/). The centroid moment tensors were obtained from the Global Centroid-Moment-Tensor Catalog (www.globalcmt.org, http://ds.iris.edu/spud/momenttensor) and National Research Institute for Earth Science and Disaster Prevention Seismic Moment Tensor Catalogue (https://www.fnet.bosai.go.jp/). Seismic wave propagation software, SES3D, used to model the waveforms is available on https://cos.ethz.ch/software/production/ses3d.html.

Acknowledgments

We express our gratitude to Göran Ekström, Andreas Zunino, Jean-Paul Montagner, Pascal Bernard, and Claudio Satriano for useful discussions and insights about this study. This work was supported by a grant from the Swiss National Supercomputing Centre (CSCS) under project ID s1040 and by the European Unions Horizon 2020 research and innovation program through an ERC Starting Grant (The Collaborative Seismic Earth Model, grant No. 714069).

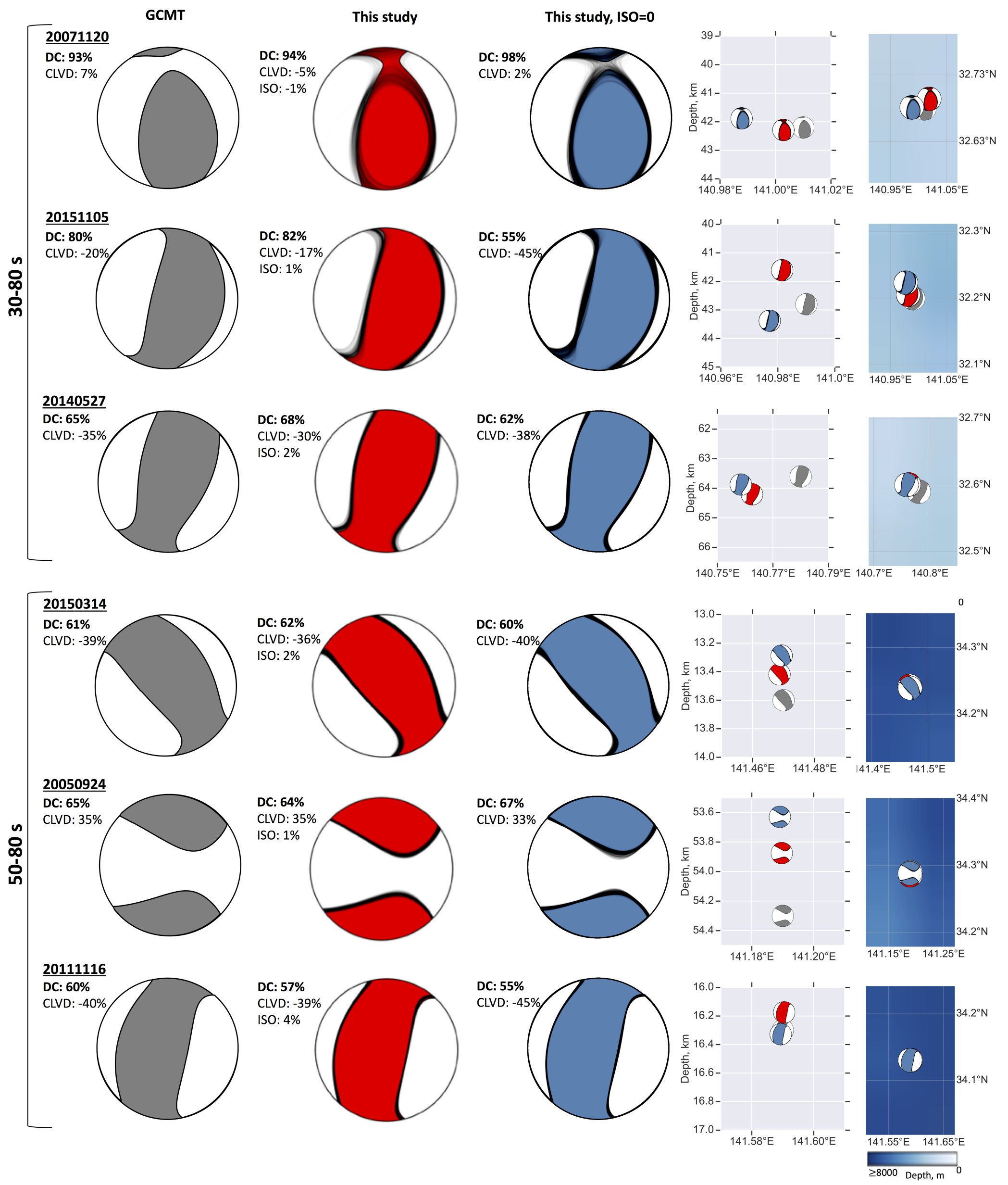

Figure 15. Moment tensor ensembles inferred with 30–80 s (top three) and 50–80 s period data (bottom three). Gray beachballs correspond to the GCMT solution, red beachball correspond to an unconstrained inversion, and blue ones correspond to the inversion, where we assume no volumetric component. Gray lines represent every 100th model of the ensemble, with the maximum-likelihood models colored. On the right, we show the maximum-likelihood locations, following the same color code.

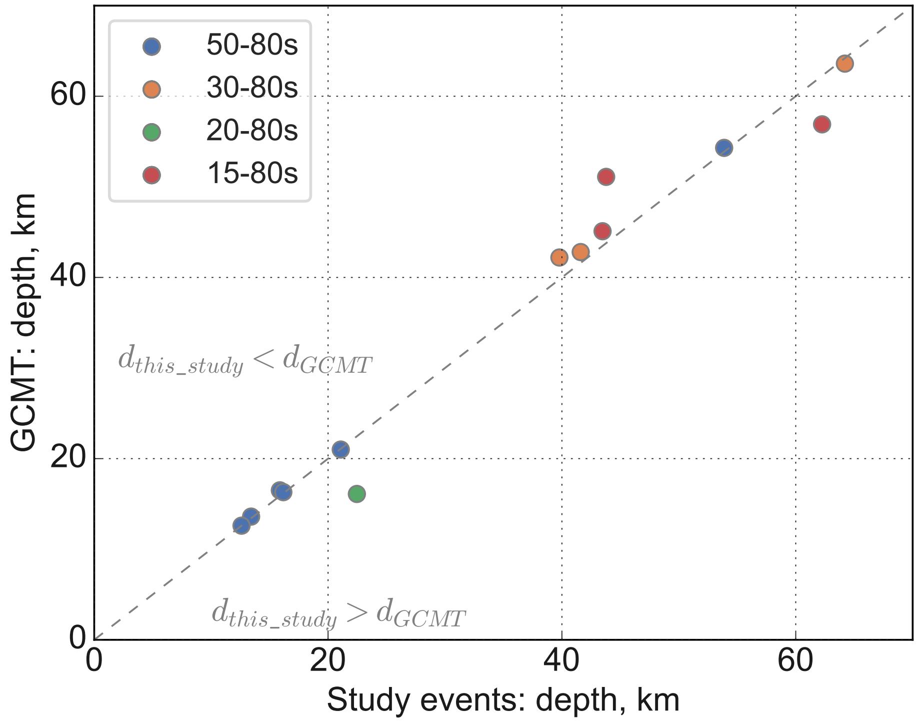

Figure 16. Comparison of earthquake depth in GCMT and unconstrained inversion of our study, color-coded by the period band of the inversion. Events inverted with the shorter-period data tend to deviate more from the GCMT solution that those inverted with longer periods.

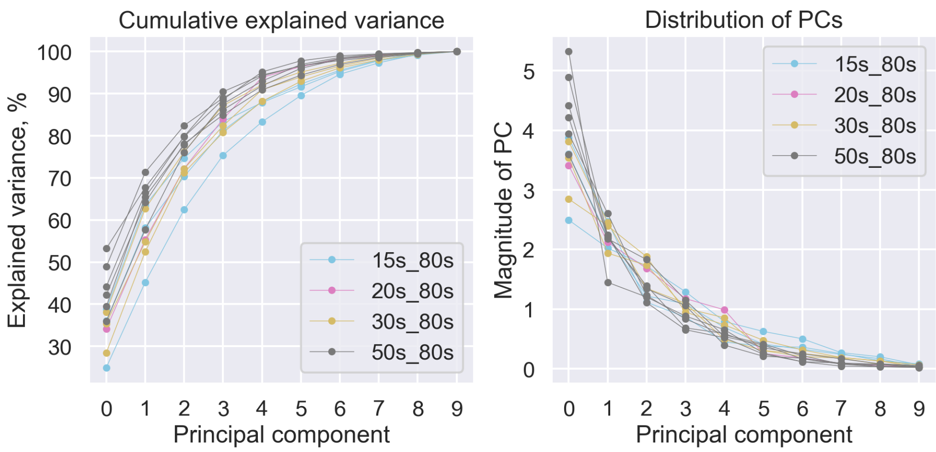

Figure 17. Left: cumulative explained variance with respect to the principal components of the posterior correlation matrix of all study events within their acceptable period band. **Right:** distribution of eigenvalues, or principal components of the posterior correlation matrix.

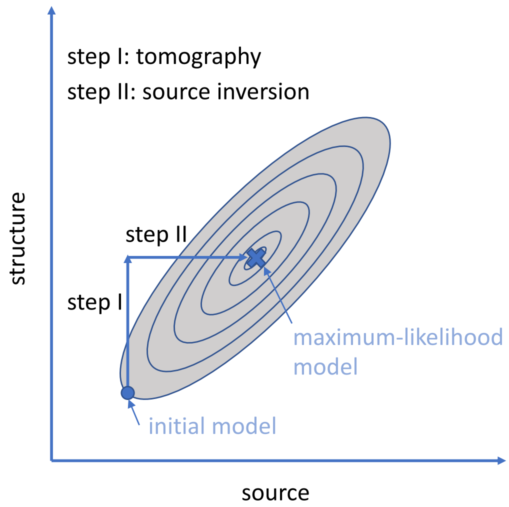

Figure 18. An idealized sketch of source and structure inversions in a two-step procedure.

References

Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D. A., … Fichtner, A. (2018, 11). Modular and flexible spectral-element waveform modelling in two and three dimensions. *Geophys. J. Int.*, *216*(3), 1675-1692. doi: 10.1093/gji/ggy469

Aki, K. (1967). Scaling law of seismic spectrum. *J. Geophys. Res. (1896-1977)*, *72*(4), 1217-1231. doi: https://doi.org/10.1029/JZ072i004p01217

Aki, K., & Richards, P. (2002). *Quantitative seismology.* University Science Books.

Alder, B. J., & Wainwright, T. E. (1959). Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.*, *31*(2), 459-466. doi: 10.1063/1.1730376

Aleardi, M., Salusti, A., & Pierini, S. (2020). Transdimensional and Hamiltonian Monte Carlo inversions of Rayleigh-wave dispersion curves: a comparison on synthetic datasets. *Near Surf. Geophys.*, *18*(5), 515-543. doi: 10.1002/nsg.12100

Babuška, V., & Cara, M. (1991). *Seismic anisotropy in the Earth.* Kluwer Academic Publishers, Dordrecht, Boston, London.

Backus, G. E. (1962). Long-wave elastic anisotropy produced by horizontal layering. *J. Geophys. Res.*, *67*, 4427-4440.

Backus, G. E., & Mulcahy, M. (1976a). Moment tensors and other phenomenological descriptions of seismic sources - I. Continuous displacements. *Geophys. J. R. Astron. Soc.*, *46*, 341-361.

Backus, G. E., & Mulcahy, M. (1976b). Moment tensors and other phenomenological descriptions of seismic sources—II. Discontinuous displacements. *Geophys. J. R. Astron. Soc.*, *47*(2), 301-329. doi: https://doi.org/10.1111/j.1365-246X.1976.tb01275.x

Bayes, T., & Price, R. (1763). An essay toward solving a problem in the doctrine of chances. *Phil. Trans. R. Soc. London*, *53*, 370-418.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [stat.ME].*

Biswas, R., & Sen, M. (2017). 2D full-waveform inversion and uncertainty estimation using the reversible jump Hamiltonian Monte Carlo. *SEG Expanded Abstracts*, 1280-1285.

Blom, N., Gokhberg, A., & Fichtner, A. (2020). Seismic waveform tomography of the central and eastern Mediterranean upper mantle. *Solid Earth*, *11*(2), 669–690. doi: 10.5194/se-11-669-2020

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., … Pugmire, D. (2016). Global adjoint tomography: First-generation model. *Geophys. J. Int.*, *207*, 1739-1766.

Burgos, G., Capdeville, Y., & Guillot, L. (2016). Homogenized moment tensor and the effect of near-field heterogeneities on nonisotropic radiation in nuclear explosion. *J. Geophys. Res. Solid Earth*, *121*(6), 4366-4389. doi: 10.1002/2015JB012744

Byerly, P. (1928). The nature of the first motion in the Chilean earthquake of November 11, 1922. *Am. J. Sci*, *16*, 232-236. doi: doi:10.2475/ajs.s5-16.93.232

Chave, A. D. (2017). Computational statistics in the earth sciences: With applications in matlab. In (pp. 281–343). Cambridge: Cambridge University Press.

Clinton, J. F., Hauksson, E., & Solanki, K. (2006, 10). An Evaluation of the SCSN Moment Tensor Solutions: Robustness of the Mw Magnitude Scale, Style of Faulting, and Automation of the Method. *Bull. Seismol. Soc. Am.*, *96*(5), 1689-1705. doi: 10.1785/0120050241

Cua, G., & Heaton, T. (2007). The virtual seismologist (vs) method: a bayesian approach to earthquake early warning. In P. Gasparini, G. Manfredi, & J. Zschau (Eds.), *Earthquake early warning systems* (pp. 97–132). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-72241-0_7

Dettmer, J., Benavente, R., Cummins, P. R., & Sambridge, M. (2014, 09). Trans-dimensional finite-fault inversion. *Geophys. J. Int.*, *199*(2), 735-751. doi: 10.1093/gji/ggu280

Dreger, D. S. (2003). 85.11 - TDMT_INV: Time Domain Seismic Moment Tensor INVersion. In W. H. Lee, H. Kanamori, P. C. Jennings, & C. Kisslinger (Eds.), *International handbook of earthquake and engineering seismology, part b* (Vol. 81, p. 1627). Academic Press. doi: https://doi.org/10.1016/S0074-6142(03)80290-5

Dreger, D. S., & Helmberger, D. (1991). Source parameters of the Sierra Madre Earthquake from regional and local body waves. *Geophys. Res. Lett.*, *18*(11), 2015-2018. doi: 10.1029/91GL02366

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B*, *195*, 216-222.

Dufumier, H., & Rivera, L. (1997). On the resolution of the isotropic component in moment tensor inversion. *Geophys, J. Int.*, *131*, 595-606.

Duputel, Z., Agram, P. S., Simons, M., Minson, S. E., & Beck, J. L. (2014, 01). Accounting for prediction uncertainty when inferring subsurface fault slip. *Geophys. J. Int.*, *197*(1), 464-482. doi: 10.1093/gji/ggt517

Duputel, Z., Rivera, L., Fukahata, Y., & Kanamori, H. (2012). Uncertainty estimations for seismic source inversions. *Geophys. J. Int.*, *190*(2), 1243-1256. doi: 10.1111/j.1365-246X.2012.05554.x

Durek, J. J., & Ekström, G. (1996). A radial model of anelasticity consistent with long-period surface wave attenuation. *Bull. Seis. Soc. Am.*, *86*, 144-158.

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. *J. Geophys. Res.*, *10*, 2825-2852.

Dziewoński, A. M., & Woodhouse, J. H. (1983). An experiment in systematic study of global seismicity: centroid-moment tensor solutions for 201 moderate and large earthquakes of 1981. *J. Geophys. Res.*, *88*, 3247-3271.

Eisner, L., & Clayton, R. W. (2001, 06). A Reciprocity Method for Multiple-Source Simulations. *Bulletin of the Seismological Society of America*, *91*(3), 553-560. Retrieved from `https://doi.org/10.1785/0120000222` doi: 10.1785/0120000222

Ekström, G., Nettles, M., & Dziewonski, A. M. (2012). The global CMT project 2004-2010: centroid moment tensors for 13,017 earthquakes. *Phys. Earth Planet. Inter.*, *200-201*, 1-9.

Engdahl, E. R., Sindorf, J. G., & Eppley, R. A. (1977). Interpretation of relative teleseismic P wave residuals. *J. Geophys. Res.(1896-1977)*, *82*(36), 5671-5682. doi: https://doi.org/10.1029/JB082i036p05671

Eshelby, J. D., & Peierls, R. E. (1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *241*(1226), 376-396. doi: 10.1098/rspa.1957.0133

Faccenna, C., Holt, A. F., Becker, T. W., Lallemand, S., & Royden, L. H. (2018). Dynamics of the Ryukyu/Izu-Bonin-Marianas double subduction system. *Tectonophysics*, *746*, 229 - 238. doi: https://doi.org/10.1016/j.tecto.2017.08.011

Ferreira, A. M. G., Weston, J., & Funning, G. J. (2011). Global compilation of interferometric synthetic aperture radar earthquake source models: 2. Effects of 3-D Earth structure. *J. Geophys. Res. Solid Earth*, *116*(B8). doi: 10.1029/2010JB008132

Ferreira, A. M. G., & Woodhouse, J. H. (2006). Long-period seismic source inversions using global tomographic models. *Geophys. J. Int.*, *166*(3), 1178-1192. doi: 10.1111/j.1365-246X.2006.03003.x

Ferreira, A. M. G., & Woodhouse, J. H. (2007, 01). Source, path and receiver effects on seismic surface waves. *Geophys. J. Int.*, *168*(1), 109-132. doi: 10.1111/j.1365-246X.2006.03092.x

Fichtner, A. (2021). *Lecture notes on inverse theory.* doi:10.33774/coe-2021-qpq2j: Cambridge Open Engage.

Fichtner, A., Bunge, H.-P., & Igel, H. (2006). The adjoint method in seismology - I. Theory. *Phys. Earth Planet. Inter.*, *157*, 86-104.

Fichtner, A., & Igel, H. (2008). Efficient numerical surface wave propagation through the optimization of discrete crustal models - a technique based on non-linear dispersion curve matching (DCM). *Geophys. J. Int.*, *173*, 519-533.

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H.-P. (2009a). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophys. J. Int.*, *179*, 1703-1725.

Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H.-P. (2009b). Spectral-element simulation and inversion of seismic waves in a spherical section of the Earth. *J. Num. An. Ind. Appl. Math.*, *4*, 11-22.

Fichtner, A., & Simutė, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. *J. Geophys. Res.*, *123*, doi:10.1002/2017JB015249.

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 Planet. Sci. Lett.*, *297*, 607-615.

Fichtner, A., van Herwaarden, D., Afanasiev, M., Simutė, S., Krischer, L., Cubuk-Sabuncu, Y., … Igel, H. (2018). The Collaborative Seismic Earth Model: Generation I. *Geophys. Res. Lett.*, *45*, 4007-4016.

Fichtner, A., Zunino, A., & Gebraad, L. (2018). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophys. J. Int.*, *2016*, doi:10.1093/gji/ggy496.

Fichtner, A., Zunino, A., Gebraad, L., & Boehm, C. (2021). Autotuning Hamiltonian Monte Carlo for efficient generalised nullspace exploration. *Geophys. J. Int.*, *227*, 941968.

Fitch, T. J., North, R. G., & Shields, M. W. (1981). Focal depths and moment tensor representations of shallow earthquakes associated with the Great Sumba Earthquake. *J. Geophys. Res. Solid Earth*, *86*(B10), 9357-9374.

Ford, S. R., Dreger, D. S., & Walter, W. R. (2009a). Identifying isotropic events using a regional moment tensor inversion. *J. Geophys. Res.*, *114*, B01306. doi: doi:10.1029/2008JB005743

Ford, S. R., Dreger, D. S., & Walter, W. R. (2009b). Source analysis of the Memorial Day explosion, Kimchaek, North Korea. *Geophys. Res. Lett.*, *36*(21). doi: https://doi.org/10.1029/2009GL040003

Fouch, M. J., & Fischer, K. M. (1998). Shear wave anisotropy in the Mariana Subduction Zone. *Geophys. Res. Lett.*, *25*(8), 1221-1224. doi: https://doi.org/10.1029/98GL00650

Frankel, A. (2013, 05). Rupture History of the 2011 M 9 Tohoku Japan Earthquake Determined from Strong-Motion and High-Rate GPS Recordings: Subevents Radiating Energy in Different Frequency Bands. *Bull. Seis. Soc. Am.*, *103*(2B), 1290-1306. doi: 10.1785/0120120148

French, S. W., & Romanowicz, B. A. (2014). Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography. *Geophys. J. Int.*, *199*, 1303-1327.

Fukuyama, E., Ishida, M., Horiuchi, S., Inoue, H., Hori, S., Sekiguchi, S., … Nonomura, K. (2001, February). *NIED Seismic Moment Tensor Catalogue January-December, 2000* (Technical Note No. 217). National Research Institute for Earth Science and Disaster Prevention.

Galitzin, B. (1909). Zur frage der bestimmung des azimuts der epizentrums eines beben. *C. R. des Seances de l'Assoc. Int. Seism. Zermatt*, 132-141.

Gebraad, L., Boehm, C., & Fichtner, A. (2020). Bayesian elastic full-waveform inversion using hamiltonian monte carlo. *J. Geophys. Res. Solid Earth*, *125*(3), e2019JB018428. doi: 10.1029/2019JB018428

Gokhberg, A., & Fichtner, A. (2016). Full-waveform inversion on heterogeneous HPC systems. *Comp. Geosci.*, *89*, 260-268.

Grandin, R., Vallée, M., Satriano, C., Lacassin, R., Klinger, Y., Simoes, M., & Bollinger, L. (2015). Rupture process of the Mw=7.9 2015 Gorkha earthquake (Nepal): Insights into Himalayan megathrust segmentation. *Geophys. Res. Lett.*, *42*(20), 8373-8382. doi: https://doi.org/10.1002/2015GL066044

Greve, S. M., & Savage, M. K. (2009). Modelling seismic anisotropy variations across the Hikurangi subduction margin, New Zealand. *Earth Planet. Sci. Lett.*, *285*(1), 16 - 26. doi: https://doi.org/10.1016/j.epsl.2009.05.035

Hallo, M., Asano, K., & Gallovič, F. (2017). Bayesian inference and interpretation of centroid moment tensors of the 2016 Kumamoto earthquake sequence, Kyushu, Japan. *Earth Planets Space*, *69*(1), 134. doi: 10.1186/s40623-017-0721-4

Hallo, M., & Gallovič, F. (2016, 08). Fast and cheap approximation of Green function uncertainty for waveform-based earthquake source inversions. *Geophys. J. Int.*, *207*(2), 1012-1029. doi: 10.1093/gji/ggw320

Hanks, T. C. (1977). Earthquake stress drops, ambient tectonic stresses and stresses that drive plate motions. *Pure Appl. Geophys.*, *115*(1), 441–458. doi: 10.1007/BF01637120

Hara, S., Fukahata, Y., & Iio, Y. (2019). P-wave first-motion polarity determination of waveform data in western Japan using deep learning. *Earth Planets Space*, *71*(1), 127. doi: 10.1186/s40623-019-1111-x

Hasegawa, A. (1990). Seismicity: Subduction zone. In *Geophysics* (pp. 1054–1061). Boston, MA: Springer US.

Hasegawa, A. (2011). Seismicity, subduction zone. In H. K. Gupta (Ed.), *Encyclopedia of solid earth geophysics* (pp. 1305–1315). Dordrecht: Springer Netherlands. doi: 10.1007/978-90-481-8702-7_14

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, *57*, 97-109.

Hayes, G. P. (2018). *Slab2 - A Comprehensive Subduction Zone Geometry Model: U.S. Geological Survey data release.* Retrieved 2020 June 8, from `https://doi.org/10.5066/F7PV6JNV`

Hayes, G. P., Rivera, L., & Kanamori, H. (2009, 09). Source Inversion of the W-Phase: Real-time Implementation and Extension to Low Magnitudes. *Seismol. Res. Lett.*, *80*(5), 817-822. doi: 10.1785/gssrl.80.5.817

Hayes, G. P., Wald, D. J., & Johnson, R. L. (2012). Slab1.0: A three-dimensional model of global subduction zone geometries. *J. Geophys. Res. Solid Earth*, *117*(B1). doi: 10.1029/2011JB008524

Hejrani, B., Tkalcic, H., & Fichtner, A. (2017). Centroid moment tensor catalogue using a 3D continental scale Earth model: application to earthquakes in Papua New Giunea and the Solomon Islands. *J. Geophys. Res.*, *122*, 5517-5543.

Helmberger, D. V., & Enge, G. R. (1980). Modeling the long-period body waves from shallow earthquakes at regional ranges. *Bull. Seis. Soc. Am.*, *70*(5), 1699-1714.

Hingee, M., Tkalčić, H., Fichtner, A., & Sambridge, M. (2011). Seismic moment tensor inversion using a 3-D structural model: applications for the Australian region. *Geophys. J. Int.*, *184*, 949-964.

Hjörleifsdóttir, V., & Ekström, G. (2010). Effects of three-dimensional Earth structure on CMT earthquake parameters. *Phys. Earth Planet. Inter.*, *179*(3), 178 - 190. doi: https://doi.org/10.1016/j.pepi.2009.11.003

Hutton, K., Woessner, J., & Hauksson, E. (2000). Earthquake Monitoring in Southern California for Seventy-Seven Years (1932-2008). *Bull. Seis. Soc. Am.*, *100*(2), 423-446. doi: doi:10.1785/0120090130

Igel, H., & Gudmundsson, O. (1997). Frequency-dependent effects on travel times and waveforms of long-period S and SS waves. *Phys. Earth. Planet. Int.*, *104*, 229-246.

Igel, H., Nissen-Meyer, T., & Jahnke, G. (2002). Wave propagation in 3D spherical sections: effects of subduction zones. *Phys. Earth Planet. Inter.*, *132*(1), 219 - 234. doi: https://doi.org/10.1016/S0031-9201(02)00053-5

Jost, M. L., & Herrmann, R. B. (1989). A student's guide to and review of moment tensors. *Seis. Res. Lett.*, *60*, 37-57.

Kanamori, H., & Given, J. W. (1981). Use of long-period surface waves for rapid determination of earthquake-source parameters. *Phys. Earth Planet. Inter.*, *27*(1), 8 - 31. doi: https://doi.org/10.1016/0031-9201(81)90083-2

Kanamori, H., & Rivera, L. (2008). Source inversion of W phase: speeding up seismic tsunami warning. *Geophys. J. Int.*, *175*(1), 222-238. doi: 10.1111/j.1365-246X.2008.03887.x

Kasahara, K. (1963). Computer program for a fault-plane solution. *Bull. Seis. Soc. Am.*, *53*(1), 1-13.

Kawakatsu, H., Kumar, P., Takei, Y., Shinohara, M., Kanazawa, T., Araki, E., & Suyehiro, K. (2009). Seismic evidence for sharp lithosphere-asthenosphere boundaries of oceanic plates. *Science*, *324*, 499-502.

Kennett, B. L. N., Engdahl, E. R., & Buland, R. (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.*, *122*, 108-124.

Khoshkholgh, S., Zunino, A., & Mosegaard, K. (2020, May). Informed Proposal Monte Carlo. *arXiv e-prints*, arXiv:2005.14398.

Knopoff, L. (1961). Analytical calculation of the fault-plane problem. In J. H. Hodgson (Ed.), *Publications of the dominion observatory* (Vol. 24, p. 309-315). Canada Department of Mines and Technical Surveys.

Koch, M. C., Fujisawa, K., & Murakami, A. (2020). Adjoint Hamiltonian Monte Carlo algorithm for the estimation of elastic modulus through the inversion of elastic wave propagation data. *Int. J. Numer. Methods Eng.*, *121*(6), 1037-1067. doi: 10.1002/nme.6256

Komatitsch, D., Erlebacher, G., Göddeke, D., & Michea, D. (2010). High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster. *J. Comp. Phys.*, *229*, 7692-7714.

Kong, X., Li, S., Wang, Y., Suo, Y., Dai, L., Géli, L., ... Wang, P. (2018). Causes of earthquake spatial distribution beneath the Izu-Bonin-Mariana Arc. *J. Asian Earth Sci.*, *151*, 90 - 100. doi: https://doi.org/10.1016/j.jseaes.2017.10.015

Krischer, L., Fichtner, A., Žukauskaitė, S., & Igel, H. (2015). Large-scale seismic inversion framework. *Seis. Res. Lett.*, *86*, 1198-1207.

Krischer, L., Igel, H., & Fichtner, A. (2018). Automated large-scale full seismic waveform inversion for North America and the North Atlantic. *J. Geophys. Res.*, *123*, doi: 10.1029/2017JB015289.

Kubo, A., Fukuyama, E., Kawai, H., & Nonomura, K. (2002). NIED seismic moment tensor catalogue for regional earthquakes around Japan: quality test and application. *Tectonophysics*, *356*(1), 23 - 48. doi: https://doi.org/10.1016/S0040-1951(02)00375-X

Kuge, K., & Kawakatsu, H. (1993). Significance of non-double couple components of deep and intermediate-depth earthquakes: implications from moment tensor inversions of long-period seismic waves. *Phys. Earth Planet. Inter.*, *75*(4), 243 - 266. doi: https://doi.org/10.1016/0031-9201(93)90004-S

Lee, E.-J., Chen, P., Jordan, T. H., & Wang, L. (2011, 07). Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California. *Geophys. J. Int.*, *186*(1), 311-330. doi: 10.1111/j.1365-246X.2011.05031.x

Lentas, K. (2017, 11). Towards routine determination of focal mechanisms obtained from first motion P-wave arrivals. *Geophys. J. Int.*, *212*(3), 1665-1686. doi: 10.1093/gji/ggx503

Li, J., Zheng, Y., Thomsen, L., Lapen, T. J., & Fang, X. (2018). Deep earthquakes in subducting slabs hosted in highly anisotropic rock fabric. *Nat. Geosci.*, *11*(9), 696–700. doi: 10.1038/s41561-018-0188-3

Liu, Q., Polet, J., Komatitsch, D., & Tromp, J. (2004). Spectral-element moment tensor inversion for earthquakes in Southern California. *Bull. Seis. Soc. Am.*, *94*, 1748-1761.

Maiti, S., & Tiwari, R. K. (2009). A Hybrid Monte Carlo Method Based Artificial Neural Networks Approach for Rock Boundaries Identification: A Case Study from the KTB Bore Hole. *Pure Appl. Geophys.*, *166*(12), 2059. doi: 10.1007/s00024-009-0533-y

Mendiguren, J. A. (1977). Inversion of surface wave data in source mechanism studies. *J. Geophys. Res. (1896-1977)*, *82*(5), 889-894. doi: 10.1029/JB082i005p00889

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.*, *21*, 1087-1092.

Minson, S. E., Simons, M., Beck, J. L., Ortega, F., Jiang, J., Owen, S. E., ... Sladen, A. (2014, 07). Bayesian inversion for finite fault earthquake source models – II: the 2011 great Tohoku-oki, Japan earthquake. *Geophys. J. Int.*, *198*(2), 922-940. doi: 10.1093/gji/ggu170

Morales-Yáñez, C., Duputel, Z., & Rivera, L. (2020, 08). Impact of 3-D Earth structure on W-phase CMT parameters. *Geophys. J. Int.*, *223*(2), 1432-1445. doi: 10.1093/gji/ggaa377

Mosegaard, K., & Sambridge, M. (2002, 04). Monte Carlo analysis of inverse problems. *Inverse Problems*, *18*, R29. doi: 10.1088/0266-5611/18/3/201

Mosegaard, K., & Tarantola, A. (2002). 16 - Probabilistic Approach to Inverse Problems. In W. H. Lee, H. Kanamori, P. C. Jennings, & C. Kisslinger (Eds.), *International Handbook of Earthquake and Engineering Seismology, Part A* (Vol. 81, p. 237 - 265). Academic Press. doi: https://doi.org/10.1016/S0074-6142(02)80219-4

Muir, J. B., & Tkalčić, H. (2020, 09). Probabilistic lowermost mantle P-Wave tomography from hierarchical Hamiltonian Monte Carlo and model parametrisation cross-validation. *Geophys. J. Int.*. (ggaa397) doi: 10.1093/gji/ggaa397

Mustać, M., & Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophys. J. Int.*, *204*, 311-323.

National Research Institute for Earth Science and Disaster Resilience. (2021). *Broadband Seismograph Network Laboratory, Network Center for Earthquake, Tsunami and Volcano.* Retrieved 8th November 2021, from `https://www.fnet.bosai.go.jp`

Neal, R. M. (1993). *Probabilistic inference using Markov chain Monte Carlo methods* (Technical Report No. CRG-TR-93-1). Dept. of Computer Science, University of Toronto.

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

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo* (chap. 5). Chapman & Hall/CRC.

Nettles, M., & Ekström, G. (1998). Faulting mechanism of anomalous earthquakes near Bárdarbunga Volcano, Iceland. *J. Geophys. Res. Solid Earth*, *103*(B8), 17973-17983.

Newrkla, K., Shiddiqi, H. A., Jerkins, A. E., Keers, H., & Ottemöller, L. (2019, 11). Implications of 3D Seismic Raytracing on Focal Mechanism Determination. *Bull. Seis. Soc. Am.*, *109*(6), 2746-2754. doi: 10.1785/0120190184

Nissen-Meyer, T., Fournier, A., & Dahlen, F. A. (2007). A two-dimensional spectral-element method for computing spherical-earth seismograms - I. Moment-tensor source. *Geophys. J. Int.*, *168*, 1067-1092.

Okamoto, T., Takenaka, H., & Nakamura, T. (2018). Evaluation of accuracy of synthetic waveforms for subduction-zone earthquakes by using a land–ocean unified 3d structure model. *Earth, Planets and Space*, *70*(1), 98. Retrieved from `https://doi.org/10.1186/s40623-018-0871-z` doi: 10.1186/s40623-018-0871-z

Omori, F. (1905). Similarity of seismic motions originating at neighboring centers. *Publ. Earthquake Invest. Com. in foreign languages (Tokyo)*, *9*(52).

Panning, M., & Romanowicz, B. (2006). A three-dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophys. J. Int.*, *167*, 361-379.

Patton, H. J., & Randall, G. E. (2002). On the causes of biased estimates of seismic moment for earthquakes in central Asia. *J. Geophys. Res. Solid Earth*, *107*(B11), 2302. doi: 10.1029/2001JB000351

Pondrelli, S., Morelli, A., Ekström, G., Mazza, S., Boschi, E., & Dziewonski, A. (2002). European–Mediterranean regional centroid-moment tensors: 1997–2000. *Phys. Earth Planet. Inter.*, *130*(1), 71 - 101. doi: https://doi.org/10.1016/S0031-9201(01)00312-0

Pugh, D. J., White, R. S., & Christie, P. A. F. (2016, 05). A Bayesian method for microseismic source inversion. *Geophys. J. Int.*, *206*(2), 1009-1038. doi: 10.1093/gji/ggw186

Romanowicz, B. A. (1982). Moment tensor inversion of long period Rayleigh waves: A new approach. *J. Geophys. Res. Solid Earth*, *87*(B7), 5395-5407. doi: 10.1029/JB087iB07p05395

Sambridge, M. S., & Gallagher, K. (2011). Inverse Theory, Monte Carlo Method. In H. K. Gupta (Ed.), *Encyclopedia of solid earth geophysics* (pp. 639–644). Dordrecht: Springer Netherlands. doi: 10.1007/978-90-481-8702-7_192

Sambridge, M. S., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Rev. Geophys.*, *40*, doi: 10.1029/2000RG000089.

Scognamiglio, L., Magnoni, F., Tinti, E., & Casarotti, E. (2016, 04). Uncertainty estimations for moment tensor inversions: the issue of the 2012 May 20 Emilia earthquake. *Geophys. J. Int.*, *206*(2), 792-806. doi: 10.1093/gji/ggw173

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

Seno, T., & Eguchi, T. (1983). Seismotectonics of the Western Pacific Region. In T. W. Hilde & S. Uyeda (Eds.), *Geodynamics of the Western Pacific-Indonesian Region* (p. 5-40). American Geophysical Union (AGU). doi: 10.1029/GD011p0005

Shang, X., & Tkalčić, H. (2020). Point-Source Inversion of Small and Moderate Earthquakes From P-wave Polarities and P/S Amplitude Ratios Within a Hierarchical Bayesian Framework: Implications for the Geysers Earthquakes. *J. Geophys. Res. Solid Earth*, *125*(2), e2019JB018492. doi: 10.1029/2019JB018492

Shuler, A., Ekström, G., & Nettles, M. (2013). Physical mechanisms for vertical-CLVD earthquakes at active volcanoes. *J. Geophys. Res. Solid Earth*, *118*(4), 1569-1586. doi: 10.1002/jgrb.50131

Siebert, L., Simkin, T., & Kimberley, P. (2010). *Volcanoes of the world, 3rd edition.* University of California Press.

Simutė, S., Steptoe, H., Gokhberg, A., & Fichtner, A. (2016). Full-waveform inversion of the Japanese islands region. *J. Geophys. Res.*, *121*, 3722-3741.

Smith, G. P., & Ekström, G. (1996). Improving teleseismic event locations using a three-dimensional Earth model. *Bull. Seis. Soc. Am.*, *86*(3), 788-796.

Staehler, S. C., & Sigloch, K. (2014). Fully probabilistic seismic source inversion - Part I: Efficient parameterisation. *Solid Earth*, *5*, 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*, doi:10.5194/se-7-1521-2016.

Swiss National Supercomputing Center. (2021). *CSCS - Swiss National Supercomputing Center.* Retrieved 8th November 2021, from `https://cscs.ch`

Takemura, S., Kimura, T., Saito, T., Kubo, H., & Shiomi, K. (2018). Moment tensor inversion of the 2016 southeast offshore Mie earthquake in the Tonankai region using a three-dimensional velocity structure model: effects of the accretionary prism and subducting oceanic plate. *Earth Planets Space*, *70*(1), 50. doi: 10.1186/s40623-018-0819-3

Takemura, S., Okuwaki, R., Kubota, T., Shiomi, K., Kimura, T., & Noda, A. (2020, 5). Centroid moment tensor inversions of offshore earthquakes using a three-dimensional velocity structure model: slip distributions on the plate boundary along the Nankai Trough. *Geophys. J. Int.*, *222*(2), 1109-1125. doi: 10.1093/gji/ggaa238

Tape, W., & Tape, C. (2013, 10). The classical model for moment tensors. *Geophysical Journal International*, *195*(3), 1701-1720. doi: 10.1093/gji/ggt302

Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. *Pure Appl. Geophys.*, *128*, 365-399.

The Global CMT Project. (2021). *Moment Tensor Product Query.* Retrieved 8th November 2021, from `https://ds.iris.edu/spud/momenttensor`

Thrastarson, S., van Driel, M., Krischer, L., Boehm, C., Afanasiev, M., van Herwaarden, D.-P., & Fichtner, A. (2020, 02). Accelerating numerical wave propagation by wavefield adapted meshes. Part II: full-waveform inversion. *Geophys. J. Int.*, *221*(3), 1591-1604. doi: 10.1093/gji/ggaa065

Thurber, C. H. (1983). Earthquake locations and three-dimensional crustal structure in the Coyote Lake Area, central California. *J. Geophys. Res. Solid Earth*, *88*(B10), 8226-8236. doi: https://doi.org/10.1029/JB088iB10p08226

Tromp, J., Tape, C., & Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophys. J. Int.*, *160*, 195–216.

Vackár, J., Burjének, J., Gallovič, F., Zahradníkk, J., & Clinton, J. (2017). Bayesian ISOLA: new tool for automated controid moment tensor inversion. *Geophys. J. Int.*, *210*, 693-705.

Vallée, M., Charléty, J., Ferreira, A. M. G., Delouis, B., & Vergoz, J. (2011). SCARDEC: a new technique for the rapid determination of seismic moment magnitude, focal mechanism and source time functions for large earthquakes using body-wave deconvolution. *Geophys. J. Int.*, *184*(1), 338-358. doi: 10.1111/j.1365-246X.2010.04836.x

Vallée, M., & Douet, V. (2016). A new database of source time functions (STFs) extracted from the SCARDEC method. *Phys. Earth Planet. Inter.*, *257*, 149 - 157. doi: https://doi.org/10.1016/j.pepi.2016.05.012

van Herwaarden, D. P., Boehm, C., Afanasiev, M., Thrastarson, S., Krischer, L., Trampert, J., & Fichtner, A. (2020, 02). Accelerated full-waveform inversion using dynamic mini-batches. *Geophys. J. Int.*, *221*(2), 1427-1438. doi: {10.1093/gji/ggaa079}

Vavryčuk, V. (2004). Inversion for anisotropy from non-double-couple components of moment tensors. *J. Geophys. Res. Solid Earth*, *109*(B7). doi: https://doi.org/10.1029/2003JB002926

Vavryčuk, V. (2015). Moment tensor decompositions revisited. *J. Seismol.*, *19*(1), 231–252. doi: 10.1007/s10950-014-9463-y

Wallace, T., Given, J., & Kanamori, H. (1982). A discrepancy between long- and short-period mechanisms of earthquakes near the long valley caldera. *Geophys. Res. Lett.*, *9*(10), 1131-1134. doi: https://doi.org/10.1029/GL009i010p01131

Wang, N., Montagner, J.-P., Fichtner, A., & Capdeville, Y. (2013). Intrinsic versus extrinsic seismic anisotropy: The radial anisotropy in reference Earth models. *Geophy Res. Lett.*, *40*, 4284-4288. doi: doi:10.1002/grl.50873

Wang, X., & Zhan, Z. (2019). Moving from 1-D to 3-D velocity model: automated waveform-based earthquake moment tensor inversion in the Los Angeles region. *Geophys. J. Int.*, *220*(1), 218-234. doi: 10.1093/gji/ggz435

Wirth, E., & Long, M. D. (2010). Frequency-dependent shear wave splitting beneath the Japan and Izu-Bonin subduction zones. *Phys. Earth Planet. Inter.*, *181*(3), 141 - 154. doi: https://doi.org/10.1016/j.pepi.2010.05.006

Woodhouse, J. H. (1983). The joint inversion of seismic wave forms for lateral variations in Earth structure and earthquake source parameter. In *Physics of the Earth's Interior* (Vol. 85, p. 366-397). Amsterdam: North-Holland.

Woodhouse, J. H., & Dziewoński, A. M. (1984). Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms. *J. Geophys. Res.*, *89*, 5953–5986.

Xu, D., & Li, D. (2008). Molecular dynamics simulation method. In D. Li (Ed.), *Encyclopedia of microfluidics and nanofluidics* (pp. 1391–1398). Boston, MA: Springer US. doi: 10.1007/978-0-387-48998-8_1052

Yu, C., Vavryčuk, V., Adamová, P., & Bohnhoff, M. (2019). Frequency-dependent moment tensors of induced microearthquakes. *Geophys. Res. Lett.*, *46*(12), 6406-6414. doi: https://doi.org/10.1029/2019GL082634

Zahradník, J., Fojtíková, L., Carvalho, J., Barros, L., Sokos, E., & Janský, J. (2015). Compromising polarity and waveform constraints in focal-mechanism solutions; the Mara Rosa 2010 Mw 4 central Brazil earthquake revisited. *J. S. Am. Earth Sci.*, *63*, 323 - 333. doi: https://doi.org/10.1016/j.jsames.2015.08.011

Zanella, G. (2020). Informed Proposals for Local MCMC in Discrete Spaces. *J. Am. Stat. Assoc.*, *115*(530), 852-865. doi: 10.1080/01621459.2019.1585255

Zhao, L., Chen, P., & Jordan, T. H. (2006). Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies. *Bull. Seis. Soc. Am.*, *96*, 1753-1763.

Zhao, L., & Helmberger, D. V. (1994). Source estimation from broadband regional seismograms. *Bull. Seis. Soc. Am.*, *84*(1), 91-104. doi: Lian-She

Zhu, L., & Helmberger, D. V. (1996). Advancement in source estimation techniques using broadband regional seismograms. *Bull. Seis. Soc. Am.*, *86*(5), 1634–1641.

Zhu, L., & Zhou, X. (2016). Seismic moment tensor inversion using 3D velocity model and its application to the 2013 Lushan earthquake sequence. *Phys. Chem. Earth*, *95*, doi:10.1016/j.pce.2016.01.002.

JOURNAL OF GEOPHYSICAL RESEARCH

Supporting Information for
"Bayesian seismic source inversion with a 3-D Earth model of the Japanese Islands"

DOI: 10.1002/

Saulė Simutė[1,2], Christian Boehm [1], Lion Krischer [1], Alexey Gokhberg [1,3],

Martin Vallée [2], Andreas Fichtner [1]

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

[2]Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

[3]Fragata Computer Systems AG, Schwyz, Switzerland

Contents of this file

1. Text S1 to S2

2. Figures S1 to S9

3. Tables S1 to S1

Introduction

In Text S1 we in more detail describe the seismic velocity model constructed for the Japanese archipelago. This model is used to compute the Green's functions for the seismic source inversion. In Text S2 we describe the procedure of choosing the mean of the prior probability density distribution, which is required for the Bayesian source inversion.

Figures S1 – S5 accompany the description of the velocity model, while figures S6 – S9 and table S1 provide more details of methodology and results of the seismic source inversion.

Text S1. The most important aspects of the full-waveform seismic velocity model building together with the summary of the results are described in the main text, section 2.

Waveform fit improvement. The overall waveform misfit decreased by 24 % after the first seven iterations in the larger initial domain, and by another 21 % during the subsequent seven iterations in the smaller domain. More details on the misfit evolution are shown in Fig. S1.

Tomographic results. The lateral averages $\bar{v}_{sv}$ and $\bar{v}_{sh}$ are shown in Fig. S2. While the depth profiles of our model largely agree with the Preliminary Reference Earth Model (PREM) model (Dziewoński & Anderson, 1981), there are a few deviations. First, the average anisotropy in the upper mantle is less pronounced in our model compared to PREM, which might be surprising given large anisotropy values throughout the model as shown in Fig. S5. The Lehmann discontinuity seen at $\sim$ 220 km depth in PREM, the origin of which is still debated (Karato, 1992; Shito et al., 2006; Revenaugh & Jordan, 1991), is not a universal feature across global models (there is no such a discontinuity in AK135 (Kennett et al., 1995)), and it is not visible in our model either. We use QL6 attenuation model (Durek & Ekström, 1996) to correct the velocities to the reference frequency of 1 Hz. The attenuation model has strong discontinuities at 80 km and 220 km depth, which can probably be attributed to the discontinuities seen at the same depth in Fig. S2.

A collection of depth and cross-sectional slices through the tomographic model in terms of deviation of the isotropic S velocity v_s from the lateral average of $\bar{v}_s$ is shown in Figs. S3 and S4. We compute isotropic S velocity as $v_s = \sqrt{\frac{2}{3}v_{sv}^2 + \frac{1}{3}v_{sh}^2}$ (e.g., Babuška & Cara, 1991; Panning & Romanowicz, 2006).

At the shallow depths (Fig. S3, 30 km) the differentiation between the thinner oceanic crust and thicker continental one is clearly visible. At 30 km depth, mantle velocities are present beneath the oceans and the Sea of Japan, while continental areas are distinguished by lower crustal velocities.

Compared to the tomographic results presented in Simutė, Steptoe, Gokhberg, and Fichtner (2016), more fine-scale features appear at depths between around 30 km and 200 km, where the sensitivity of the fundamental mode surface waves is highest for the periods of 15–80 s (Dahlen & Tromp, 1998) (Figs. S3, S4). The Pacific plate slab at the Japan trench and the Philippine Sea plate slab at the Nankai trough, characterized as positive velocity anomalies, become more pronounced at 100–150 km depth. Negative velocity anomalies associated with the volcanic arc in central Japan, become more localized at shallow depths ($<$ 100 km), while the negative velocity anomaly beneath the Shikoku basin grows stronger in the north, just off the Kii peninsula (Fig. S3, 150 km depth). More fine-scale structure appears beneath Ulleung Island in the Sea of Japan, with strong negative velocity anomalies extending down to $\sim$ 200 km depth (Fig. S4), below which it resembles a narrow tail, well visible at a cross section at latitude 37° (Fig. S4).

While we chose the perturbations of the isotropic S wave velocity in order to give a general view of the velocity anomalies, we note that anisotropy is required to fit the waveforms, and hence we show depth slices of anisotropy computed as $\zeta = \frac{v_{sh} - v_{sv}}{v_s} \times 100$ %. Although, as expected, it is dominated by positive values, meaning v_{sh} is larger than v_{sv}, at depths of 80 km and deeper, areas of negative anisotropy arise (Fig. S5).

Text S2.

In this section we describe the procedure for choosing the mean of the prior probability density for the Bayesian source inversion.

We run a preliminary inversion for all the study events at periods of 15–80 s. For this initial sampling we use automatically selected measurement windows over the main energy part of the trace, use the Global Centroid-Moment-Tensor (GCMT) solution (The Global CMT Project, 2021) as a prior mean, and run the inversion for 5×10^4 samples. If the synthetic data for the maximum-likelihood solution provide a significantly better waveform fit than the GCMT solution, we use it as a prior mean for the main inversion, both for constrained and unconstrained scenarios.

For two out of three events, for which the 15–80 s period band proved to be the acceptable inversion period band (20040407, 20050816), the prior mean model is updated. We perform this preliminary step at all period-band inversions for these two events. Hence, at each inversion period band, these events have different priors (Fig. S6).

For events which could not be inverted at the shortest-period band, we always use the GCMT solution as a prior mean. Omitting the preliminary inversion step was done due to the time constraints. In short, for all the events with an acceptable period band other than 15–80 s, the GCMT solution is used as a prior mean.

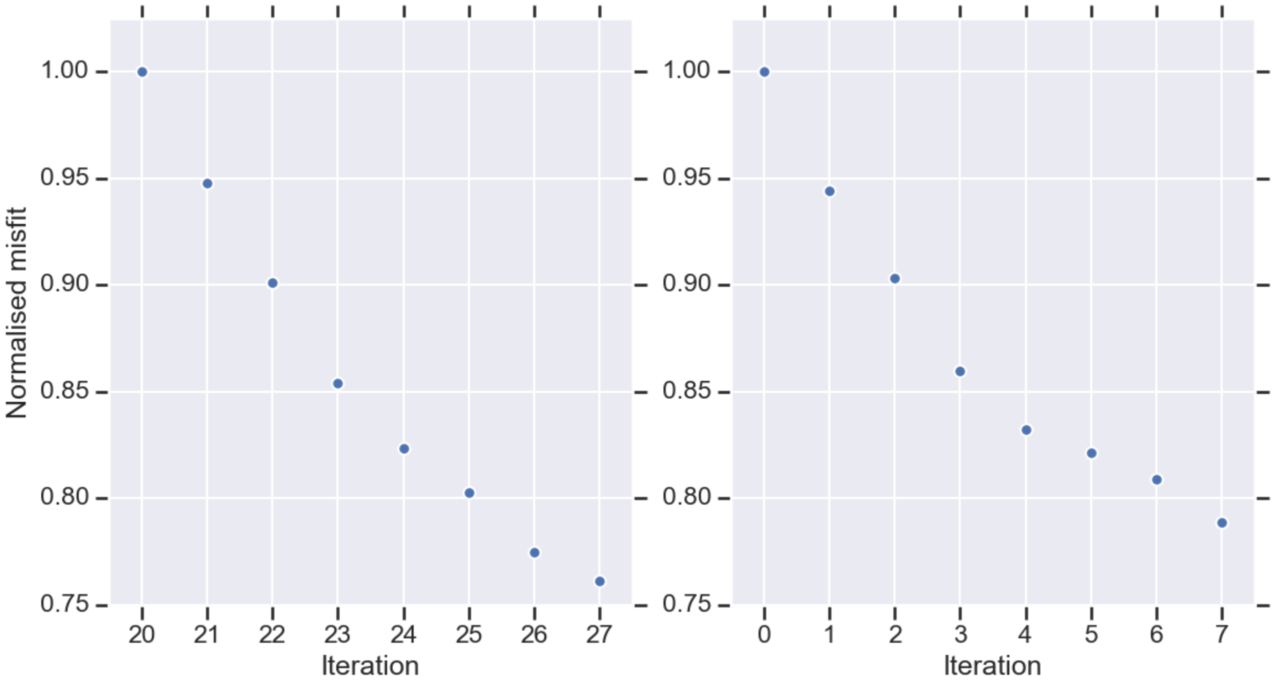

Figure S1. Per-iteration misfit reduction for measurement windows in the period band of 15–80 s, computed for time-frequency misfit as used in the inversion. **Left:** waveform fit improved by 24 % in the first seven iterations run with the original model setup. **Right:** waveform fit improved further by 21 % after seven iterations run with the new model setup as shown in Fig. 1.

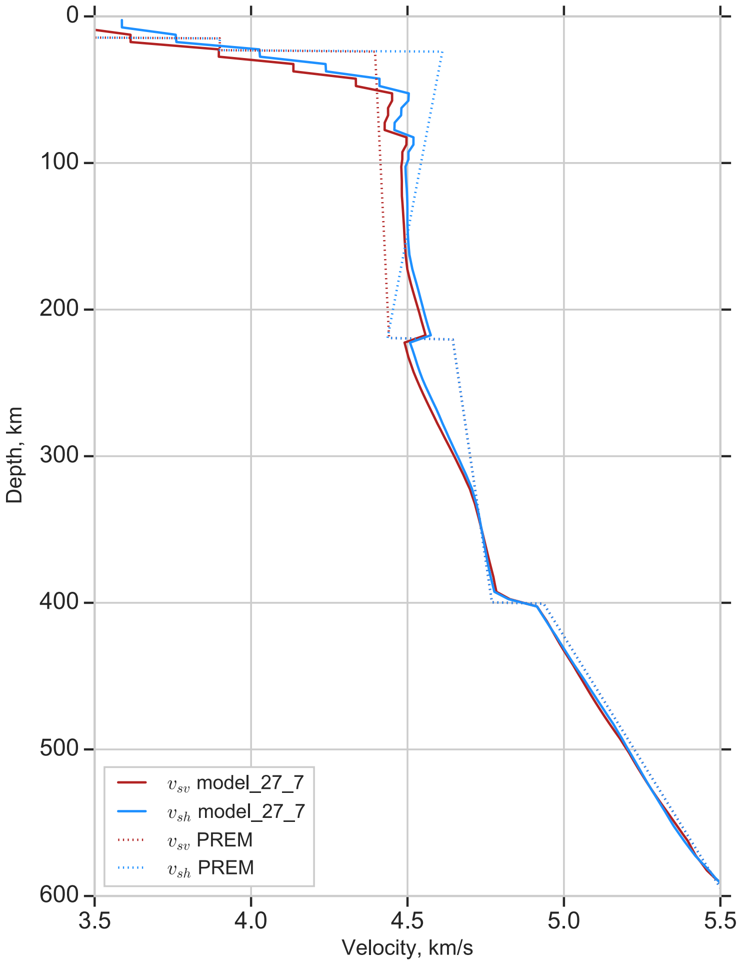

Figure S2. Lateral averages of the updated model of v_{sv} and v_{sh} at a frequency of 1 Hz. The equivalent velocities for the radially symmetric PREM are shown for comparison (Dziewoński & Anderson, 1981).

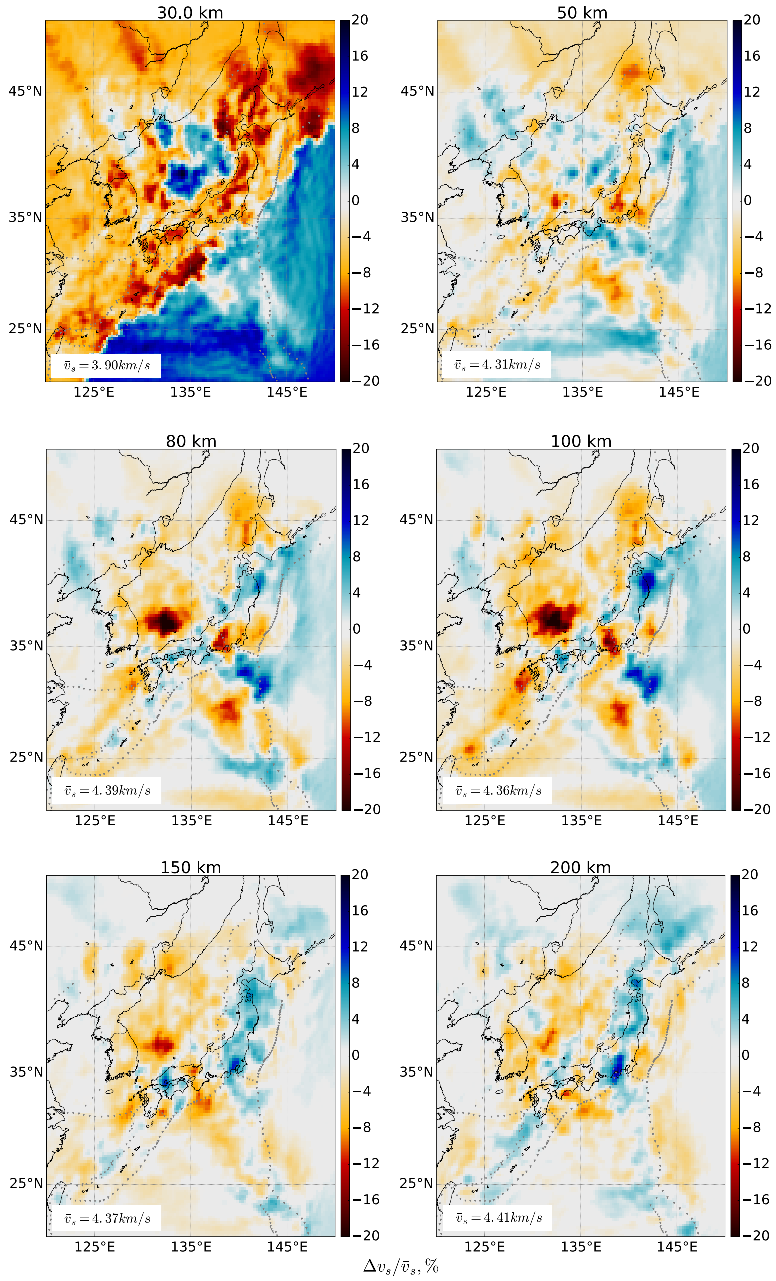

Figure S3. Depth slices of isotropic v_s perturbations across the model domain. Perturbation is computed as $\frac{v_s - \bar{v}_s}{\bar{v}_s}$ $\times 100$ %, where $\bar{v}_s$ is the lateral average of v_s for each depth. At 30 km depth we can see the mantle velocities in the Pacific ocean as well as the Sea of Japan. The negative velocity anomaly in the Sea of Japan is strongest in amplitude and extent between 80 and 100 km depths. Positive velocity anomalies are associated with the subducting Pacific and Philippine Sea slabs. Dashed gray lines represent plate boundaries.

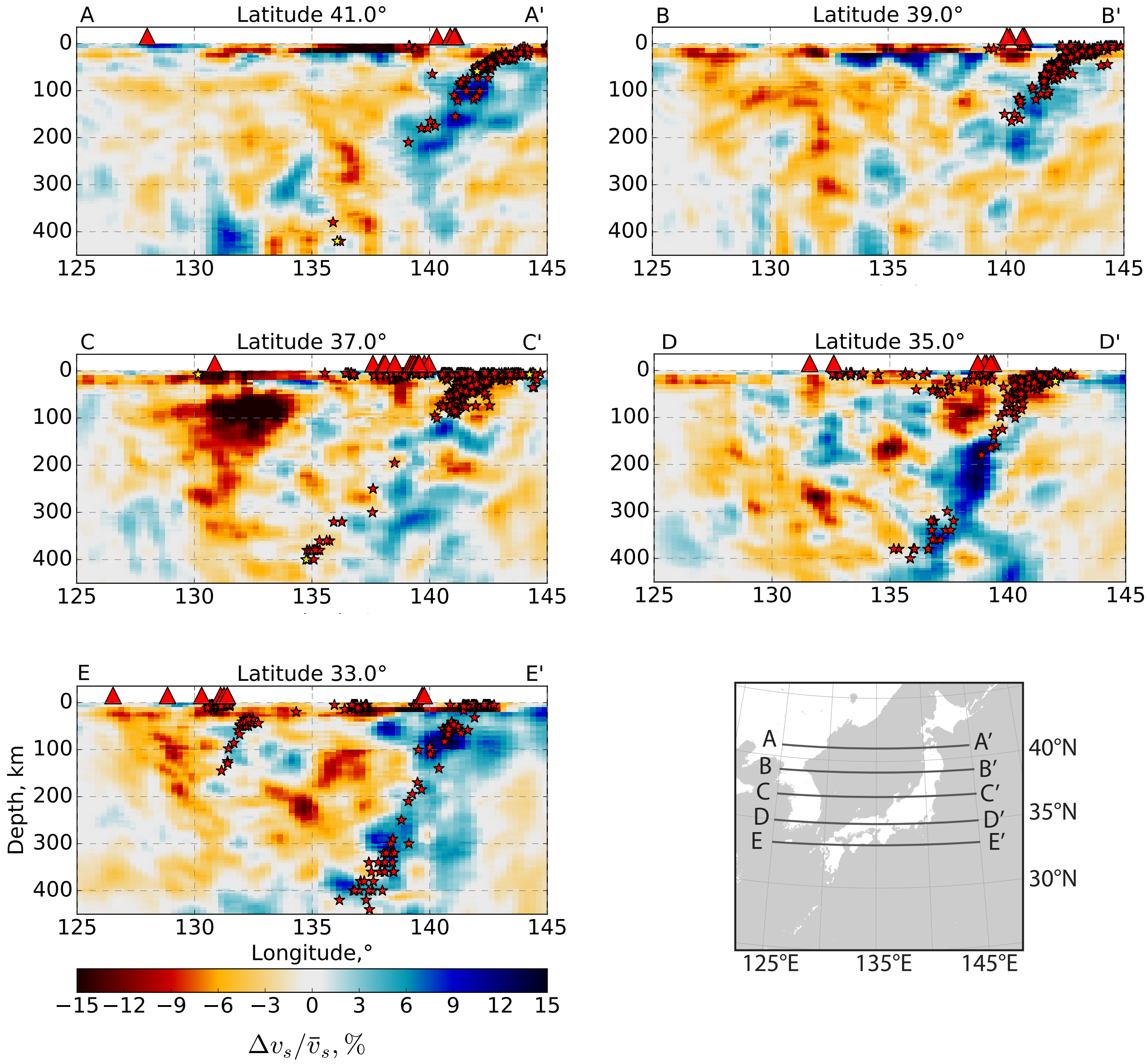

Figure S4. Cross-sections through the isotropic v_s perturbations in the model domain. Red and yellow stars represent earthquakes since 1997 and earthquakes used in this study, respectively, within 1° of the slice. Red triangles represent Holocene volcanoes (Siebert et al., 2010).

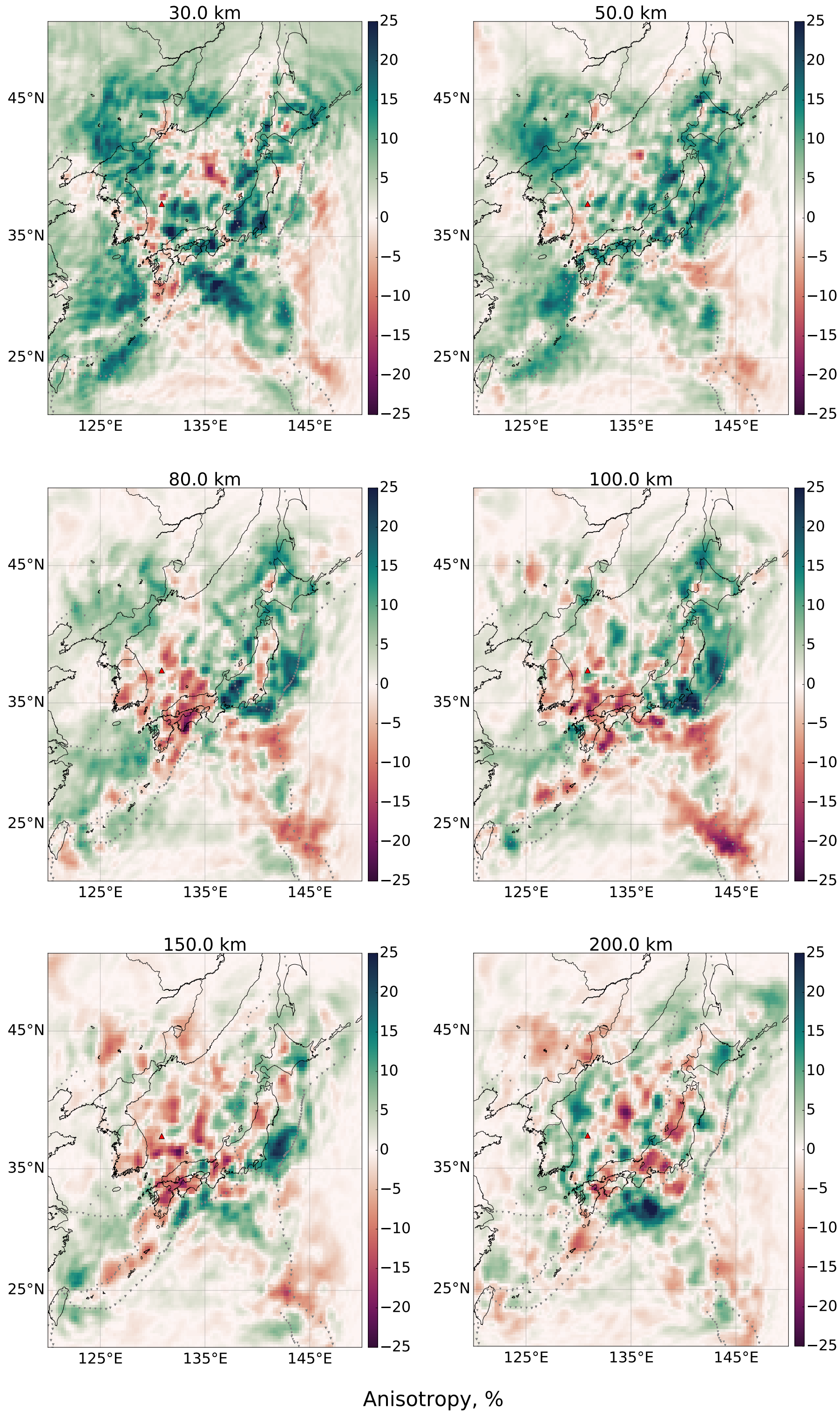

Figure S5. Radial anisotropy $\zeta = \frac{v_{sh}-v_{sv}}{v_s} \times 100\%$ for selected depth slices across the model domain.

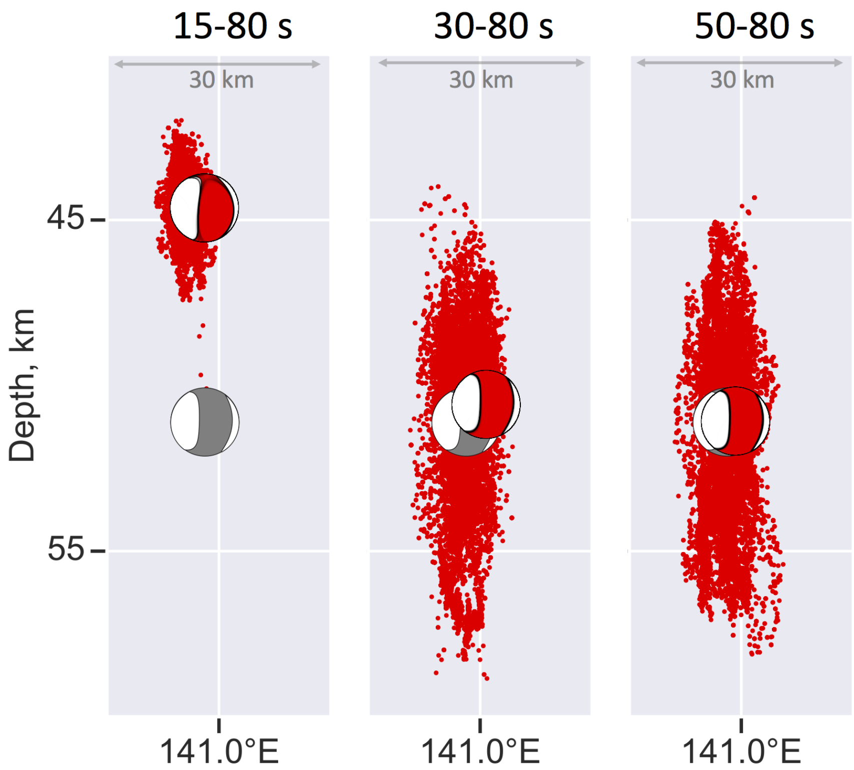

Figure S6. Ensembles of the location for event 20050816 after the preliminary inversions, performed in order to update the location used as a prior mean in the main inversions. The results are for three period bands, i.e., 15–80 s, 30–80 s, and 50–80 s. Red dots represent the accepted models from the unconstrained inversions. The maximum-likelihood models from each inversion are shown as red focal mechanisms, and they are used as prior means in the main inversions. Gray beachballs represent the GCMT solution, which was used as a prior mean in each of these preliminary runs.

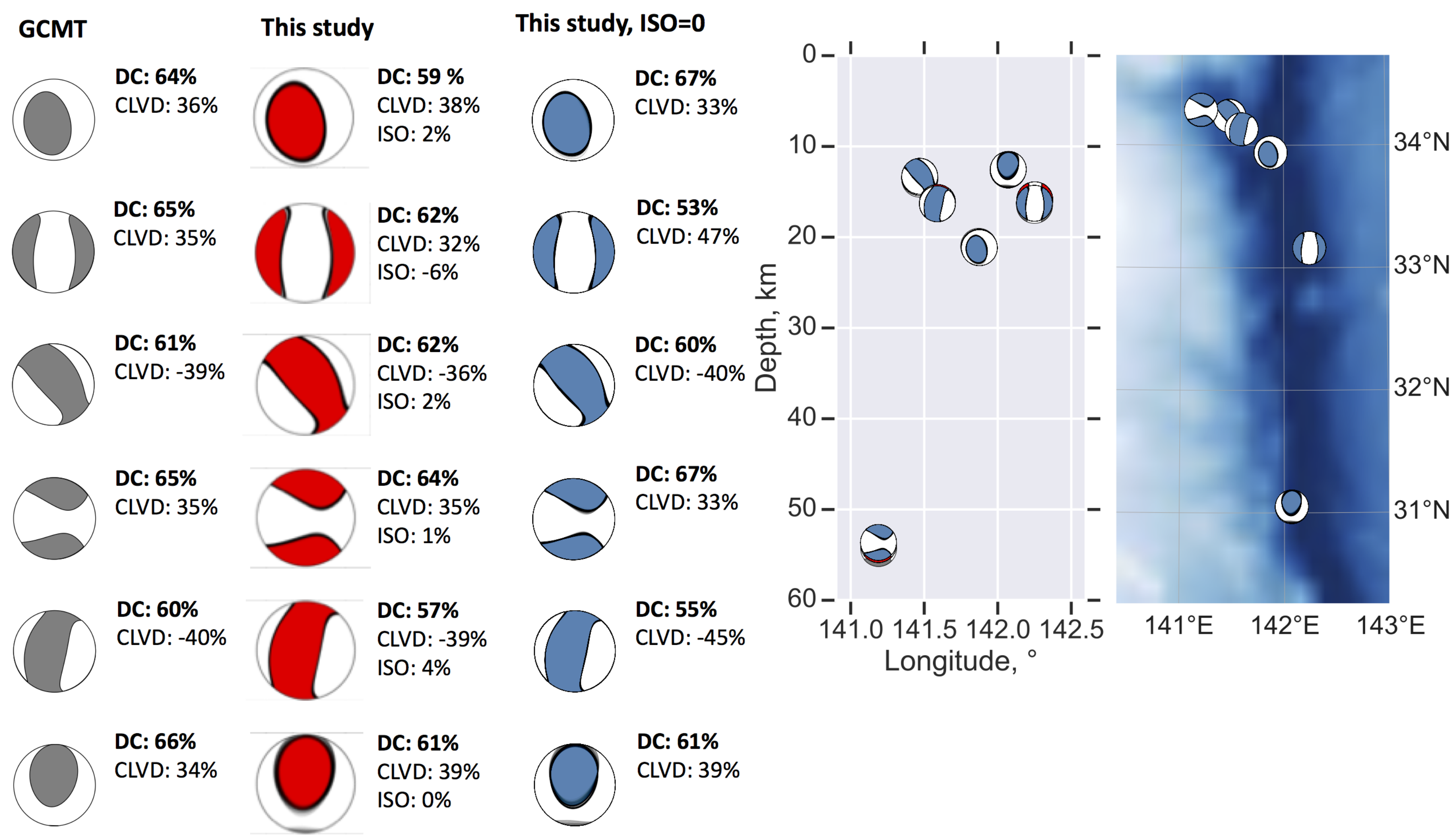

Figure S7. Moment tensor ensembles inferred with 50–80 s period data (events in the third, fourth and fifth rows are also presented in the main text). Gray beachballs correspond to the GCMT solution, red beachballs correspond to an unconstrained inversion, and blue ones correspond to the inversion, where we assume no volumetric component. Gray lines represent every 100th model of the ensemble, with the maximum-likelihood models colored. On the right, we show the maximum-likelihood locations, following the same color code.

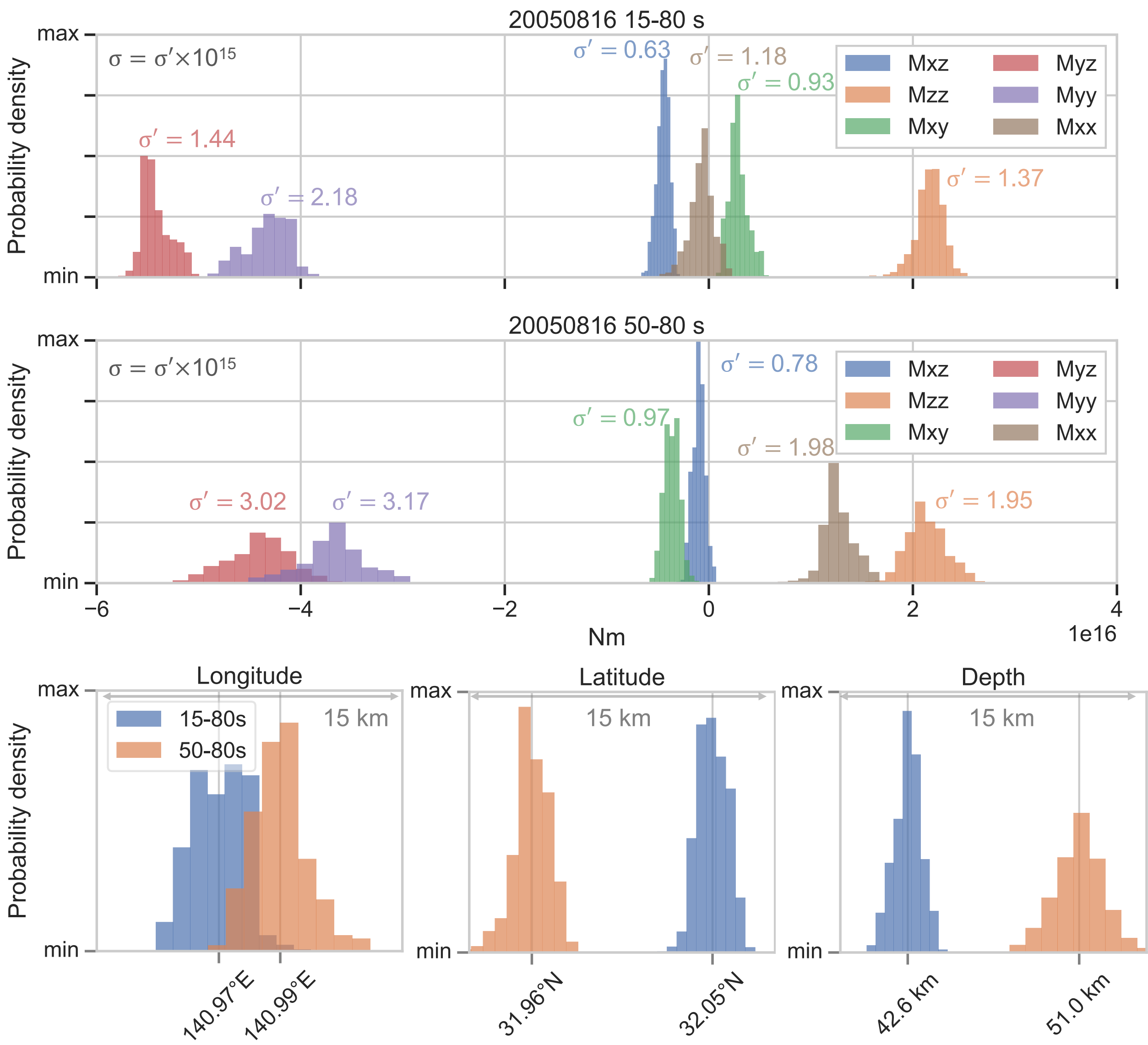

Figure S8. Comparison of marginal probability densities inferred with 15–80 s and 50–80 s data for event 20050816. Marginal probability densities for moment tensor elements with standard deviation values are shown in the top two graphs. Marginal probability densities for location parameters are shown below. Generally, shorter-period inversion better constrains the inversion parameters, especially the diagonal elements (M_{xx}, M_{yy}, M_{zz}) of the moment tensor and depth.

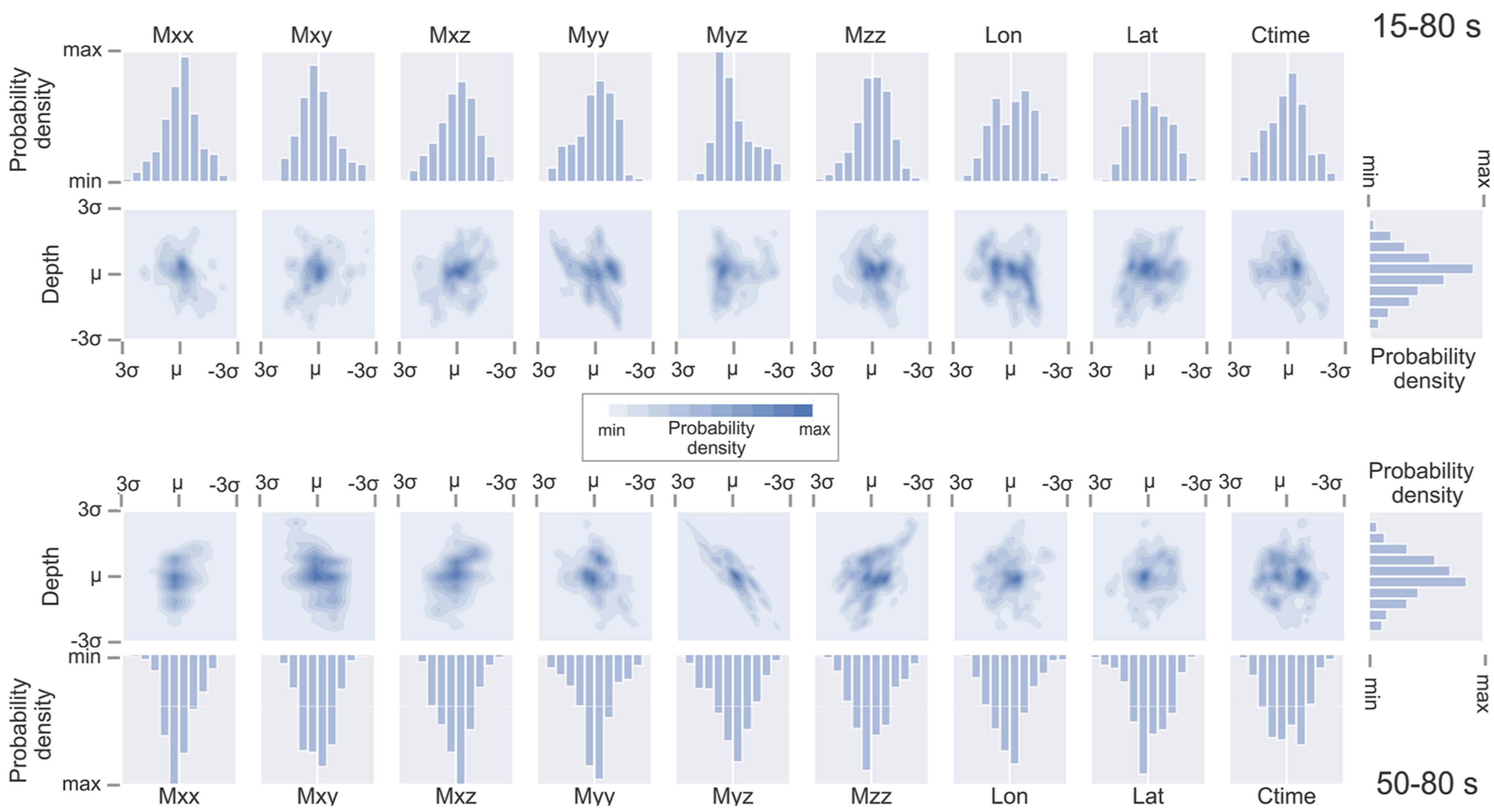

Figure S9. Comparison of selected trade-offs and marginal probability density functions for 15–80 s (top) and 50–80 s (bottom) period data inversion of event 20050816. The limits for each parameter depend on the corresponding standard deviations, and are set from $\mu - 3\sigma$ to $\mu + 3\sigma$, where μ is the mean and σ is the standard deviation of the distributions. At long periods, depth has considerable trade-offs with M_{yz} and M_{zz} components, but the parameters are much better constrained when short period data are used.

Event ID	Date	Origin time	Lon	Lat	Depth, km	Seismic moment, 1e16 Nm	CLVD, %
15 – 80 s							
20170506	2017-05-06	14:23:17.4	140.78	33.21	56.9	3.488	-52
20040407	2004-04-07	13:47:02.7	140.50	34.27	45.1	4.836	-62
20050816	2005-08-16	19:15:34.2	140.98	31.96	51.1	5.606	-47
20 – 80 s							
20121126	2012-11-26	20:08:00.8	141.9	33.43	16.1	2.049	48
30 – 80 s							
20151105	2015-11-05	08:23:03.5	140.99	32.20	42.8	5.388	-20
20071120	2007-11-20	11:46:31.4	141.01	32.68	42.2	3.787	7
20140527	2014-05-27	16:12:10.5	140.78	32.59	63.6	2.837	-35
50 – 80 s							
20101129	2010-11-29	15:02:11.0	141.87	33.94	21.0	2.690	36
20050204	2005-02-04	18:34:13.7	142.25	33.16	16.5	3.831	35
20150314	2015-03-14	15:36:08.8	141.47	34.24	13.6	4.811	-39
20050924	2005-09-24	14:48:48.3	141.19	34.29	54.3	7.49	35
20111116	2011-11-16	00:43:58.1	141.59	34.13	16.3	9.73	-40
20150508	2015-05-08	12:11:29.6	142.07	31.05	12.6	2.370	34

Table S1. Information on the study events, as given in the GCMT catalog (The Global CMT Project, 2021). Earthquakes are grouped according to their shortest acceptable inversion period band.

References

Babuška, V., & Cara, M. (1991). *Seismic anisotropy in the Earth.* Kluwer Academic Publishers, Dordrecht, Boston, London.

Dahlen, F. A., & Tromp, J. (1998). *Theoretical Global Seismology.* Princeton University Press. NJ.

Durek, J. J., & Ekström, G. (1996). A radial model of anelasticity consistent with long-period surface wave attenuation. *Bull. Seis. Soc. Am.*, *86*, 144-158.

Dziewoński, A. M., & Anderson, D. L. (1981). Preliminary reference Earth model. *Phys. Earth Planet. Inter.*, *25*, 297–356.

Karato, S.-I. (1992). On the Lehmann discontinuity. *Geophysical Research Letters*, *19*(22), 2255-2258. doi: https://doi.org/10.1029/92GL02603

Kennett, B. L. N., Engdahl, E. R., & Buland, R. (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.*, *122*, 108-124.

Panning, M., & Romanowicz, B. (2006). A three-dimensional radially anisotropic model of shear velocity in the whole mantle. *Geophys. J. Int.*, *167*, 361-379.

Revenaugh, J., & Jordan, T. H. (1991). Mantle layering from ScS reverberations: 3. The upper mantle. *J. Geophys. Res. Solid Earth*, *96*(B12), 19781-19810. doi: https://doi.org/10.1029/91JB01487

Shito, A., Karato, S., Nishihara, Y., & Matsukage, K. (2006). Towards mapping the tree-dimensional distribution of water in the upper mantle from velocity and attenuation tomography. In S. D. Jacobsen & S. van der Lee (Eds.), *Earth's deep water cycle, geofigure 14. summary of seismological observations and theoretical and experimental results with respect to rs/p. b03308 shito et al.: Attenuation and velocity in izu-bonin 12 of 13 b03308 earth's* *deep water cycle, geophys. monogr. ser* (Vol. 168, p. 225-236). AGU, Washington, D. C.

Siebert, L., Simkin, T., & Kimberley, P. (2010). *Volcanoes of the world, 3rd edition.* University of California Press.

Simutė, S., Steptoe, H., Gokhberg, A., & Fichtner, A. (2016). Full-waveform inversion of the Japanese islands region. *J. Geophys. Res.*, *121*, 3722-3741.

The Global CMT Project. (2021). *Moment Tensor Product Query.* Retrieved 8th November 2021, from `https://ds.iris.edu/spud/momenttensor`