Lightweight joint inversion of point-source moment-tensor and station-specific time shifts

Thanh-Son Phạm[1]

[1]Research School of Earth Sciences, The Australian National University, Canberra, ACT, Australia

Key Points:

- The presence of station-specific time shifts largely due to 3D Earth structures is a practical challenge in regional moment tensor inversion
- Existing joint inversion methods of moment tensor parameters and time shifts are computationally expensive
- This study proposes a lightweight joint inversion method utilizing modern infrastructures available in deep learning frameworks

Corresponding author: T.-S. Phạm, ThanhSon.Pham@anu.edu.au

Abstract

The misalignment of the observation and predicted waveforms in regional moment tensor inversion is mainly due to seismic models' incomplete representation of the Earth's heterogeneities. Current moment tensor inversion techniques, allowing station-specific time shifts to account for the model error, are computationally expensive. Here, we propose a lightweight method to jointly invert moment-tensor parameters and unknown station-specific time shifts utilizing the modern functionalities in deep learning frameworks. A L_2^2 misfit function between predicted synthetic and time-shifted observed seismograms is defined in the spectral domain, which is differentiable to all unknowns. The inverse problem is solved by minimizing the misfit function with a gradient descent algorithm. The method's feasibility, robustness, and scalability are demonstrated on earthquakes in the Long Valley Caldera, California. This work presents an example of fresh opportunities to apply advanced computational infrastructures developed in deep learning to geophysical problems.

Plain Language Summary

Understanding the physics of a seismic source is critical to discriminating its nature, which can be of tectonic or volcanic origins or artificial explosions. However, the major challenge in practice is the incomplete knowledge of the Earth's structures to explain observed seismological records adequately. In many cases, the imperfect knowledge of the Earth's structures can be well approximated by allowing the prediction to be time-shifted to match the observation waveform. But, the shifting amounts are unknown, which introduces a new set of unknowns to solve apart from the parameterization of the seismic source. Here, we utilize recent advancements in scientific computation from a deep learning toolbox to develop an effective solver for the moment tensor inversion problem. We demonstrate that our method can robustly yield the seismic source solution with much less computational effort than existing methods. This presents an exciting opportunity to tackle geophysical problems with computationally powerful tools developed in other communities.

1 Introduction

A seismic moment tensor (MT) is a mathematical representation of a seismic source under the point-source assumption in space and time for small-to-moderate earthquakes when using relatively long-period waveforms (Aki & Richards, 2002). A full MT is a 3×3 matrix, but for most underground earthquakes, the net torque is negligible, and the matrix is symmetric with six unknown elements (e.g., Jost & Herrmann, 1989; Stein & Wysession, 2003). In this study, we use the north-east-down coordinate system for an MT, denoted as,

$$\mathbf{m} = [m_k; k = 1 \dots 6] = [m_{nn}, m_{ee}, m_{dd}, m_{ne}, m_{nd}, m_{ed}] \in \mathcal{R}^6. \tag{1}$$

In MT inversion, a set of Green functions, or Earth's structure responses, $\mathbf{G}^k(\boldsymbol{\xi}, \mathbf{x}, t)$, is computed in advance for six orthogonal bases of the MT space. We assume the source, $\boldsymbol{\xi}$, and receiver $\mathbf{x}$ locations are known in the scope of this study. Then, a tensor (linear) multiplication predicts the observed waveforms,

$$\mathbf{d}(t) = \mathbf{m} \cdot \mathbf{G}(\boldsymbol{\xi}, \mathbf{x}, t) = \sum_k m_k \mathbf{G}^k. \tag{2}$$

$\mathbf{G} \in \mathcal{R}^{n_r \times n_c \times 6 \times n_t}$ is a forth-order tensor and $\mathbf{d} \in \mathcal{R}^{n_r \times n_c \times n_t}$ is a third-order tensor, where n_r, n_c and n_t are number of receivers, number of components in each receiver and number of samples, respectively. The MT solution, $\mathbf{m}^*$, can be approximated by its least-square estimate of the linear system,

$$\mathbf{m}^* = \mathbf{d} \cdot \mathbf{G}^T \cdot (\mathbf{G} \cdot \mathbf{G}^T)^{-1}. \tag{3}$$

The superscript ^{T} denotes the matrix transposition.

However, in practice, the numerical predictions and observed waveforms are often not aligned (Figs. 1A, B), and the misalignment could severely affect the MT solution estimate (Zhu & Helmberger, 1996; Zhao & Helmberger, 1994). Let τ be an unknown station-specific time shift to align the observation to the prediction, and Equation 2 can be re-written as,

$$\mathbf{d}(t-\tau) = \mathbf{m} \cdot \mathbf{G}(\boldsymbol{\xi}, \mathbf{x}, t). \tag{4}$$

The first source of misalignment is due to the unknown origin time of the inverted seismic source. This leads to an imprecise window of the observed waveforms to be matched with the synthetic waveforms. The imprecise origin time lead to a baseline time shift applied to all stations' waveforms.

Secondly, the lack of complete knowledge of the Earth's structures results in path-specific time shifts (Zhu & Helmberger, 1996; Zhao & Helmberger, 1994; Zhu & Ben-Zion, 2013; Silwal & Tape, 2016). Indeed, a 1D velocity model is often used in regional surface wave inversion. If the model is faster or slower than the real Earth along the path toward a receiver, the observed waveform needs to be shifted backward or forward to account for the structural error. The presence of unknown station-specific time shifts makes the MT inversion strongly nonlinear, whose existing computational methods are computationally expensive.

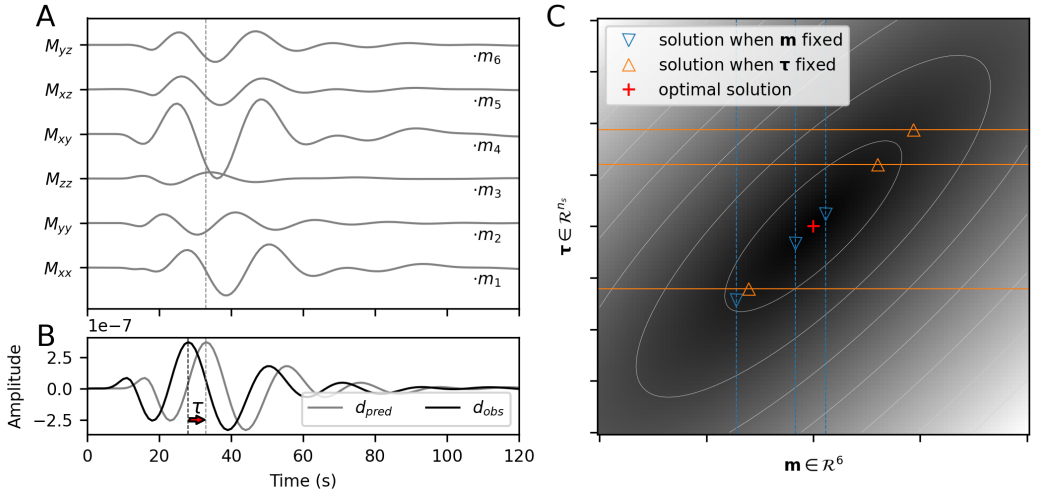


Figure 1. Motivating problem for joint inversion of moment tensor (MT) parameters, $\mathbf{m}$, and unknown station-specific time shifts, $\boldsymbol{\tau}$. (A) The waveforms are Green functions corresponding to elementary moment tensors, $M_i, i = 1, \ldots, 6$. (B) The typical mismatch between observed and predicted waveforms. The observed waveform, d_{obs}, needs to be shifted by an unknown time shift, τ, to match the predicted waveform, d_{pred}. (C) Schematic demonstration of a global minimum in the joint parameter space of MT parameters and station-specific time shifts, $\mathbf{m}$ and $\boldsymbol{\tau}$. Horizontal and vertical colored lines represent two widely-used approaches for MT inversion with unknown time shifts. The vertical lines sketch the idea behind the cut-and-paste method (Zhao & Helmberger, 1994; Zhu & Helmberger, 1996), while the horizontal lines sketch the idea behind the time domain moment tensor method (D. Dreger et al., 2000). Here we consider full MT solutions, so $\mathbf{m} \in \mathcal{R}^6$, and one time shift for each station, so $\boldsymbol{\tau} \in \mathcal{R}^{n_r}$, where n_r is the number of seismic stations.

Indeed, the cut-and-paste (CAP) algorithm (Zhu & Helmberger, 1996; Zhao & Helmberger, 1994) and its variation (Zhu & Ben-Zion, 2013) are widely used for regional full-waveform MT inversion where time-shifts are allowed to account for model errors. In this algorithm, a misfit value is computed for an imputed MT solution assuming the optimal alignment between predicted and observed waveforms can be determined by cross-correlation (depicted by vertical blue lines in Fig. 1c). Then, the best MT solution is grid-searched as the smallest misfit over the entire parameter MT space, which often results in a costly computational burden. For example, Silwal and Tape (2016) and Alvizuri et al. (2018) recently implemented this approach for full MT inversion using a uniform parameterization of MT space (Tape & Tape, 2015). The grid search was implemented on a cluster due to the computational scale (Alvizuri et al., 2018; Thurin et al., 2022) given that 20 million trials were evaluated for each earthquake.

Alternatively, an MT solution can be solved efficiently using least-square estimation given an imputed set of station-specific time shifts (depicted by horizontal orange lines in Fig. 1c). After marching the grid of all possible combinations of station-specific time shifts, the optimal MT solution corresponds to the minimum misfit (e.g., D. Dreger et al., 2000). More recently, Hejrani et al. (2017) assumed a single time shift for all stations to account for possible errors in the cataloged centroid time. The grid search for station-specific time shifts is also costly due to the exponentially larger number of combinations to explore, especially when the analysis involves a larger number (i.e., more than 10) of stations.

The ensemble sampling method in the Bayesian framework in the joint space of MT parameters and unknown time shifts (Viltres et al., 2022; Hu et al., 2023) is a relatively independent approach of the above-mentioned approaches. This approach has recently become possible thanks to the advance of powerful probabilistic sampling methods (Goodman & Weare, 2010; Roberts & Rosenthal, 2001; Del Moral et al., 2006). It narrows the targeted sampling near the optimal solution, which provides a useful estimate of the parameters' uncertainty. However, the sampling approach suffers a severe scalability problem because the sampler becomes inefficient given more seismic stations. If n_r stations are used simultaneously in the inversion and a time shift is assumed for each station, the dimension of the unknown space, $6+n_r$, grows, while the MT parameters occupy just a small subspace.

In this study, we propose a lightweight method in the optimization framework to tackle the nonlinear inverse problem. First, we represent the misfit function between the prediction and time-shifted observation in the spectral domain, where the misfit is differentiable to all dependent parameters. Consequently, the misfit function can be minimized using a gradient descent algorithm to obtain an optimal solution. The algorithm is implemented in *TensorFlow*, a popular deep-learning framework, to utilize its advanced auto-differentiation functionality and computational performance. The method's feasibility, robustness, and scalability are bench-marked against real data from several Long Valley Caldera volcanic events.

2 Methodology

2.1 Data preparation and pre-processing

Seismic waveforms and station metadata were downloaded from the North California Earthquake Data Center (NCEDC) using their web service. The retrieved waveforms were corrected for instrumental responses for velocity seismograms, then filtered between 20–50 s, with an acausal, 4-corner Butterworth bandpass filter, and down-sampled to 1 sample per second. We visually inspected the pre-processed waveforms and eliminated those having glitches or exhibiting anomalously high noise levels. They were then arranged into a 3-dimensional tensor in $\mathcal{R}^{n_r \times n_c \times n_t}$ where $n_r, n_c = 3$, and n_t are the

numbers of receivers, channels, and time samples. In this study, all waveforms are 250 s long (i.e., $n_t = 250$) starting from the origin time reported in the NCEDC catalog.

2.2 Calculation of Green's functions

Synthetic velocity seismograms of an impulse source corresponding to six orthogonal MT bases were generated using the frequency-wave number method in the Computer Program in Seismology package (Herrmann, 2013). We used the 1D South California model (D. S. Dreger & Helmberger, 1990) like previous studies of earthquakes in the region (D. Dreger et al., 2000; Minson & Dreger, 2008). The calculated seismograms are filtered and re-sampled in the same way as the observed data and arranged into a 4-dimensional tensor $\mathbf{G} \in \mathcal{R}^{n_s \times n_c \times n_e \times n_t}$ and $n_e = 6$ is the number of MT orthogonal bases.

2.3 Differentiable misfit function and optimized inversion

Without losing the generality, we consider the inverse problem with an unknown time shift, τ, for a single observed seismogram. An L_2^2 misfit function between predicted and shifted (or translated) observed waveforms is defined as,

$$J(\mathbf{m}, \tau) = \int_t |\mathbf{m} \cdot \mathbf{G}(t) - \mathbf{d}(t - \tau)|^2 \, dt. \tag{5}$$

The integration over time is a notional representation of the summation over discrete digital waveform samples.

A time-shifted function in the time domain is a multiplication in the frequency domain, f, via a Fourier transform,

$$\mathcal{F}\{\mathbf{d}(t - \tau)\}(f) = \mathcal{F}\{\mathbf{d}(t)\}(f) \ \exp(-i\omega\tau), \tag{6}$$

where $\omega = 2\pi f$. The Fourier transform of the integral kernel in Equation 5 becomes,

$$\mathcal{F}\{\mathbf{m} \cdot \mathbf{G}(t) - \mathbf{d}(t - \tau)\}(f) = \mathcal{F}\{\mathbf{m} \cdot \mathbf{G}(t)\}(f) - \mathcal{F}\{\mathbf{d}(t)\}(f) \ \exp(-i\omega\tau). \tag{7}$$

For convenience, we denote $\hat{\mathbf{G}} := \mathcal{F}\{\mathbf{G}(t)\}(f)$, $\hat{\mathbf{d}} := \mathcal{F}\{\mathbf{d}(t)\}(f)$. Because the L_2^2 norms in spectral and temporal domains equal (i.e., Parseval's theorem — Gradshteyn et al., 2000), the misfit function, Equation 5, can be re-written in the spectral domain as,

$$J(\mathbf{m}, \tau) = \int_f |\mathbf{m} \cdot \hat{\mathbf{G}} - \hat{\mathbf{d}} \ \exp(-i\omega\tau)|^2 \, df. \tag{8}$$

In Equation 8, $|\cdot|$ is the absolute value of the complex spectra.

The cost function, $J(\mathbf{m}, \tau)$, defined in Equation 8 is continuous and differentiable to all dependent variables, $\mathbf{m}$ and τ, indeed,

$$\frac{\partial J}{\partial \mathbf{m}} = 2 \int_f (\hat{\mathbf{G}}_r \cdot \hat{\mathbf{D}}_r + \hat{\mathbf{G}}_i \cdot \hat{\mathbf{D}}_i) \, df, \tag{9}$$

and

$$\frac{\partial J}{\partial \tau} = 2\omega \int_f (\hat{\mathbf{D}}_r \cdot (\hat{\mathbf{d}}_i \cos\omega\tau - \hat{\mathbf{d}}_r \sin\omega\tau) + \hat{\mathbf{D}}_i \cdot (\hat{\mathbf{d}}_i \cos\omega\tau + \hat{\mathbf{d}}_r \sin\omega\tau)) \, df, \tag{10}$$

where

$$\mathbf{D}_r = (\mathbf{m} \cdot \hat{\mathbf{G}}_r - \hat{\mathbf{d}}_r \cos\omega\tau - \hat{\mathbf{d}}_i \sin\omega\tau)$$
$$\mathbf{D}_i = (\mathbf{m} \cdot \hat{\mathbf{G}}_i - \hat{\mathbf{d}}_i \cos\omega\tau + \hat{\mathbf{d}}_r \sin\omega\tau).$$

$\mathbf{D}$ denotes the difference of the spectra of the observed and predicted waveforms in the integral kernel in Equation 8. Subscripts i, r denote the real and imaginary components

of the Fourier transforms of Green's tensor and observed data. With the differentiable misfit function, a gradient descent algorithm can be used to derive the optimum corresponding to the inverse solution in the joint space of MT parameters and time shift.

In the remainder of this work, we assume a single time shift for all three components of one station to demonstrate the method's feasibility. By assuming so, we ignore the Earth's structure anisotropy, which could result in different time shifts in surface wave arrivals in radial and tangential seismograms. The time shifts are proxy to account for isotropic heterogeneity in 3D Earth structures when approximated by 1D velocity models (Zhu & Helmberger, 1996; Zhao & Helmberger, 1994; Zhu & Ben-Zion, 2013). Thus, our inversion has n_r time shifts and 6 MT parameters as unknowns.

2.4 Algorithm implementation

Our implementation utilizes the auto differentiation functionality available on TensorFlow (*www.tensorflow.org*), a popular deep learning framework. The partial derivatives, Equations 9 and 10, can be automatically calculated by back-propagating the forward definition of the cost function, Equation 8. The framework also offers well-tested, high-performance gradient descent algorithms, such as Adam (Kingma & Ba, 2014). In the present implementation, we use the Adam optimizer as a default option due to its empirical performance. As demonstrated later, the implementation enables acceleration on a graphic processing unit (GPU), resulting in excellent scalability to data size without a negligible increase in compute time. As we utilize the core functions of a modern Deep Learning framework, the inversion algorithm can also be implemented in other frameworks with similar functionalities, such as PyTouch, as of the developers' preference.

Because the event's magnitude is generally unknown, the amplitudes of observed seismograms and pre-computed Green's functions can be in different dynamic ranges. This is an inherent challenge for iterative algorithms because the initial solution must be relatively close to the optimum to guarantee the algorithm's convergence. Here, we normalize the observation and Green tensor with their absolute medians to mitigate the magnitude scaling difference. The actual magnitude of the MT solution will be restored afterward using the ratio between the median values. Empirically, we found a learning rate of 0.2 (or others in the same magnitude order) performs well for the examples presented in this paper. The learning rate defines the relative step length of the descending iterations.

3 Results

3.1 Mw 4.9, 1997/11/22 17:20:35, Long Valley Caldera volcanic event

Firstly, we analyze data from the Long Valley Caldera event (Mw 4.9, 1997/11/22 17:20:35) to demonstrate the feasibility and robustness of the method proposed in this study. This event, often referred to as LV2, was in a sequence of large events during the volcanic unrest from 1997–1999 (D. Dreger et al., 2000; Minson & Dreger, 2008; Phạm & Tkalčić, 2021). We used the hypo-center reported in the NCEDC catalog with a depth of 5.1 km to calculate the synthetic Green tensor before the inversion. All 15 broadband seismic stations of the Berkeley Digital Seismograph Network found within the searching box (Fig. 2A) were used in this experiment.

Fig. 2 summarises the inversion result. As shown in Fig. 2C, the cost function descends quickly in less than 100 epochs (i.e. iteration steps) before stabilizing in the subsequent epochs. The optimal solution of MT parameters and station-specific time shifts was selected to yield the smallest cost in the iteration chain. The optimal MT solution has dominant 52% double couple (DC), large isotropic (ISO: 33%), and compensated lin-

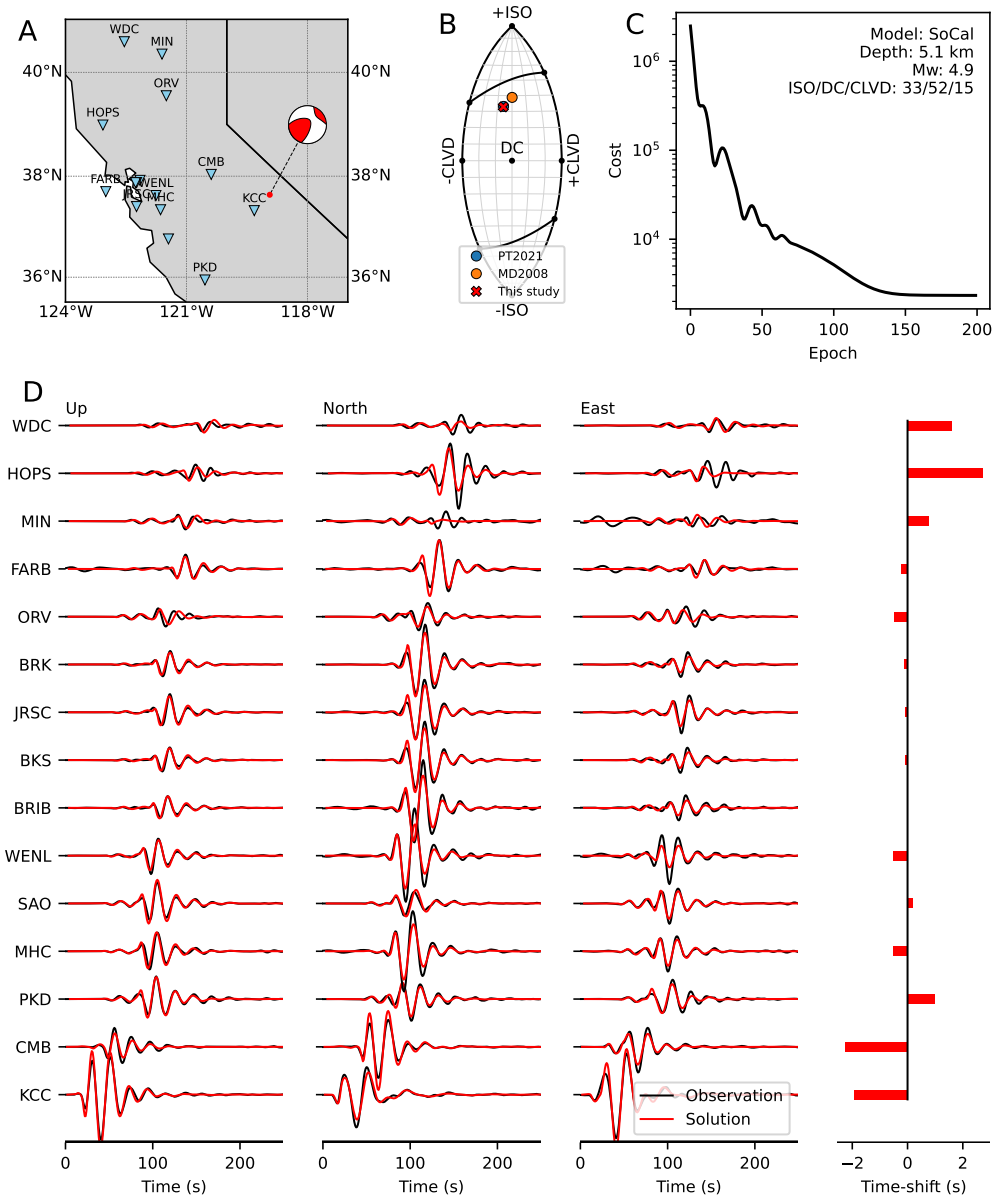


Figure 2. Summary of MT solution for the Mw 4.9, 1997/11/22 17:20:35, Long Valley Caldera volcanic event, referred to as LV2 in Minson and Dreger (2008) and Phạm and Tkalčić (2021). (A) Location map of the event and monitoring stations. The beach ball shows the deviatoric moment tensor solution of this study. (B) The source type of MT solution is denoted on a lune diagram of source types (Tape & Tape, 2012). (C) Evolution of descending cost (Equation 8) as a function of epochs. The text denotes the used Earth model (D. S. Dreger & Helmberger, 1990), earthquake depth from North California Earthquake Data Center, the solution's moment magnitude and its standard decomposition into isotropic (ISO), double couple (DC), and compensated linear vector dipole (CLVD) components. (D) Observed data and best-predicted waveforms are shown in black and red, respectively. The right panel shows the output station-specific time shifts.

ear dipole (CLVD: 15%) components (Fig. 2B). Predicted waveforms corresponding to the solution exhibit a good fit with observed waveforms after time-shifted (Fig. 2D).

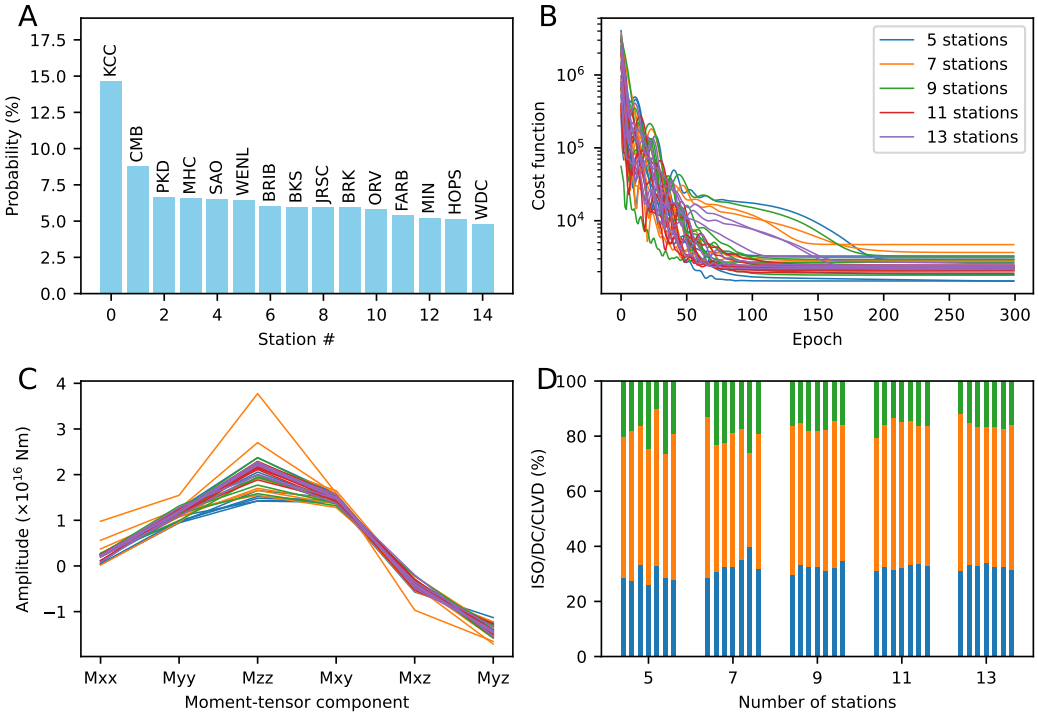


Figure 3. Demonstration of solution robustness for random selections of seismic stations. (A) Probability of selecting a station as a function of epicentral distance to the earthquake location. (b) Cost functions descend as a function of epochs for multiple runs. (c) Optimal MT solutions of the runs, color-coded as a number of stations involved. (d). ISO/CLVD/DC component decomposition corresponds to the optimal solutions.

To further examine the robustness of the recovered solution, we repeat the MT inversion for multiple subsets of 5, 7, 9, 11, and 13 stations drawn from the original set of 15. The inversion was repeated seven times for each of the subset sizes. In each run, a seismic station was selected randomly until the subset size was reached. The probability of picking a station is inversely proportional to its distance to the earthquake source (Fig. 3A), or in other words, a station closer to the source is more likely to be picked. All inversions ran for 300 epochs, although all cost functions descended to convergence within about 200 epochs (Figs. 3B). The collection of results shows the similarity element-wise in all optimal MT solutions, given that the largest variability was encountered when 7 stations were used (Fig. 3C). Despite the variation, the standard MT decomposition reveals a consistent CLVD component in the solution at around 18%.

The presence of the significant CLVD component in the MT solution of this event (Figs. 2B) confirms the findings by Phạm and Tkalčić (2021). It is worth noting that Phạm and Tkalčić (2021) and this study both considered the Earth's structural error despite the differences in the approaches. The previous study considered the influence of the 1D velocity model's error by simulating synthetic waveforms and their covariance matrices. This study, however, utilizes station-specific time shifts as a proxy to tolerate Earth model error, which largely results in waveform mismatch. The agreement of a significant CLVD component for this event, which was overlooked in the past (Phạm & Tkalčić,

2021), revealed by two different approaches, corroborates the importance of properly treating the Earth model's error in MT inversion procedures.

3.2 Robustness of recovered station-specific time shifts

While the previous subsection features the robustness of resolving MT parameters, this section demonstrates the robustness of the recovered time shifts. We retrieve data from six events in Long Valley Caldera's 1997–1999 unrest sequence to do so. A set of 15 stations recording the events are used for the inversion to compare relative variation in time shifts (Fig. 4a).

Fig. 4 summarises the consistency in the recovered time shifts. The selected events show a wide variety of source types. Four events with significant isotropic components are likely related to volcanic activities, while the other two with dominant DC components are possibly of pure tectonic nature (Figs. 4A and B). Despite the diversity in source types, the recovered station-specific time shifts are consistent among all six events (Fig. 4c). The anomalous time shift of station HOPS in event 1997/11/30 21:17:05 is due to noisy data waveforms.

Overall, the Southern California model (D. S. Dreger & Helmberger, 1990) provides a good 1D Earth model for the LVC events as needed time shifts for most stations are consistently small. Nevertheless, it appears consistently slower along paths to close stations, such as KCC or SAO, but slightly faster along the paths to more distant stations, such as PKD and WDC.

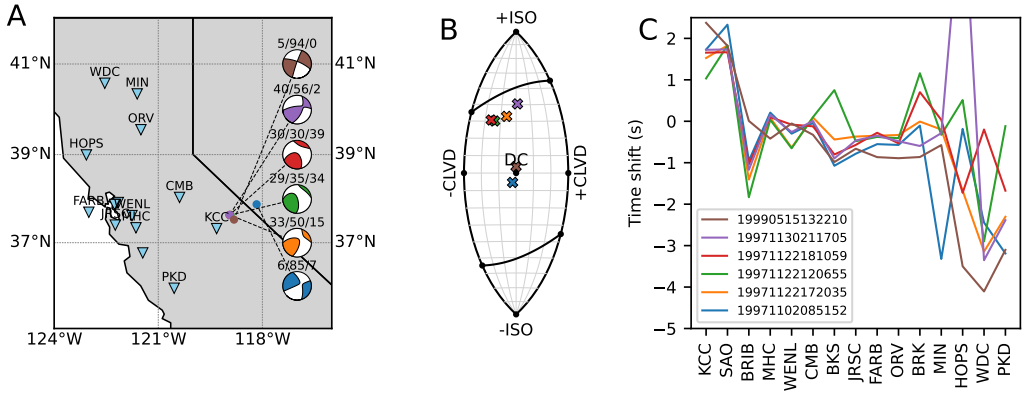


Figure 4. Demonstration of robustness in recovered station-specific time shifts. (A) A map of six events in and near the Long Valley Caldera region and 15 stations was used in the experiment. (B) Recovered source types of the events plotted on a lune diagram (Tape & Tape, 2012). (C) Relative station-specific time shifts recovered for the six events.

3.3 Performance scalability with number of stations

On another important aspect, we show the proposed method's excellent scalability to tens of seismic stations. This experiment utilizes 49 stations recording the Mw xx 2021/07/09 01:46:00 event in the LVC region (Fig. 5A). The MT inversions were repeated for data from increasing numbers of stations with a step of two. To compare the running time, we terminated each run after 200 epochs without considering the actual convergence of the MT solutions (Fig. 5B).

When running on a single Central Processing Unit (CPU) of a compute node on the National Computing Infrastructure's Gadi cluster, the compute time only doubled when the input size was increased by 10 times, exhibiting the CPU code's strong sub-linear scalability. The computing time is even faster on the cluster's Graphical Processing Unit (GPU). Yet more significantly, the input size increase does not apparently affect compute time. This starkly contrasts with existing methods, as discussed in the Introduction, whose efficacy depends greatly on the input sizes.

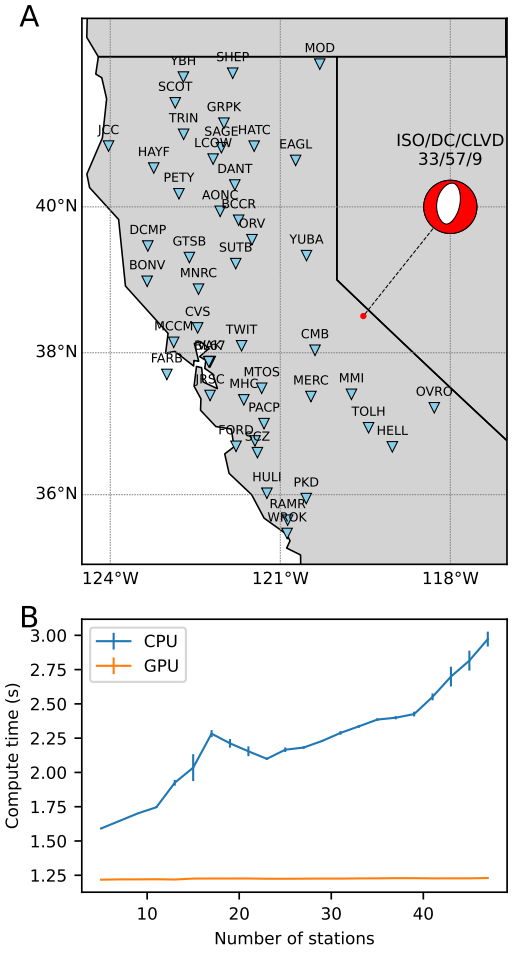


Figure 5. Demonstration of excellent computational scalability. (A) Location map of event and seismic stations recording event Mw xx 2021/07/09 01:46:00. (B) Colored lines show the computing times on a Central Processing Unit (CPU) and a Graphical Processing Unit (GPU) of the National Computing Infrastructure's Gadi cluster as functions of the number of stations involved in the inversion.

4 Discussion and Conclusion

In the previous section, we demonstrated a lightweight method's feasibility, robustness, and scalability to invert MT parameters and station-specific time shifts jointly. Here we discuss some aspects of the proposed algorithm, which can be further explored in future research.

In the current implementation, the source location is yet to be considered unknown to be constrained by data because the waveform derivatives with respect to source location are not available in our forward waveform simulation. However, the current algorithm, which is lightweight, can be repeated efficiently in a grid search surrounding a preliminary source location (Hejrani et al., 2017; Mustać & Tkalčić, 2016). Translating the existing forward methods (e.g., Herrmann, 2013) into a modern programming framework where auto-differentiation functionality is supported, e.g., TensorFlow or PyTouch, could be a promising way to incorporate the source location into the present inversion framework.

It is worth noting that the point-wise L_2^2 misfit function (Equations 5 and 8) assumes uncorrelated data noise in the observation. Recent studies demonstrate the benefit of considering correlated noise characterized by symmetric covariance matrices, which can be estimated either from ambient noise (Mustać & Tkalčić, 2016; Vackář et al., 2017; Duputel et al., 2012) or involve the prior assumptions of Earth's structural error (Phạm & Tkalčić, 2021; Vasyura-Bathke et al., 2021). Incorporating the covariance matrices into this study's proposed algorithm is practical, but its robustness is subject to further investigation. Similarly, it is feasible, but yet to be verified, to incorporate an unknown source time function (STF), representing the history of realizing energy rather than an impulse source in time, when the STF is represented as a weighted linear combination of several bases (Stähler & Sigloch, 2014).

Because the proposed joint inversion is cast in the optimization framework, it is not benefited significantly if using uniform MT parameterization methods (Tape & Tape, 2015; Stähler & Sigloch, 2014) due to their non-linearity. Instead, we chose the primitive MT parameterization for the present implementation, so the derivatives to MT parameters can be calculated rapidly (Equation 9). For the same reason, we expect the implementation's performance to be similar when using other linear MT parameterization methods (Kikuchi & Kanamori, 1982; Kawakatsu, 1996), which are convenient when full MT is not required, for example, deviatoric MT (Hejrani & Tkalčić, 2018; Hejrani et al., 2017).

The use of L_2^2 misfit function between predicted and observed waveform could pose a concern regarding the presence of local minima, most likely due to the waveform cycle-skipping. Thanks to the proposed solution's agility, one quick solution is to repeat the inversion with multiple initializations surrounding the empirically guessed time shift configuration. Using novel waveform similarity measurements, which are not subjected to local minima, such as the Warsterstein distance (Sambridge et al., 2022), could be an interesting future research topic for this application.

The main limitation of the presented method in an optimization framework is the lack of sufficient uncertainty estimates concerning inverted solutions because it only outputs the optimal solution. This is of critical importance when applying for highly non-unique settings such as critically shallow earthquakes or explosions (Hejrani & Tkalčić, 2020; Kawakatsu, 1996; Ford et al., 2010; Alvizuri et al., 2018; Hu et al., 2023). However, thanks to the availability of the derivatives, it is readily available for further investigation of effective gradient-based Hamiltonian Monte Carlo sampler (Fichtner & Simutė, 2018).

In conclusion, we present a lightweight inversion method in an optimization framework for a classical problem in regional MT inversion, the joint inversion of MT parameters, and station-specific time shifts. The station-specific time shifts can be considered a simplified proxy for 3D Earth structure errors with incomplete Green's functions in MT inversion problems. In doing so, the L_2^2 misfit function between the prediction and shifted observation is cast in the frequency domain thanks to Parseval's theorem. Tests on pilot events from the Long Valley Calderas demonstrated the proposed method's feasibility, robustness, and scalability. This paper hopes to highlight a fresh opportunity

to benefit from the computational infrastructures thanks to the rapidly growing artificial intelligence communities for geophysical problems.

Data and code availability

Data for this study come from the Berkeley Digital Seismic Network (BDSN), doi:10.7932/BDSN, operated by the UC Berkeley Seismological Laboratory, which is archived at the Northern California Earthquake Data Center (NCEDC), doi:10.7932/NCEDC.

Codes and data examples of the pilot event, as shown in Fig. 2, can be found at https://github.com/tsonpham/JointMTS.git.

Acknowledgments
This work greatly benefited from discussion with Jinyin Hu and Hrvoje Tkalčić through ongoing research on moment tensor inversion. The Department of Defence Air Force Research Laboratory's grant, contract number FA9453-20-C-0072, supported the author's postdoc at The Australian National University. He also acknowledges financial support from the Australian Research Council through a Discovery Early Career Researcher Award, project DE230100025. This research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), supported by the Australian Government.

References

Aki, K., & Richards, P. G. (2002). *Quantitative Seismology* (Second edition ed.). Sausalito, CA: University Science Books.

Alvizuri, C., Silwal, V., Krischer, L., & Tape, C. (2018). Estimation of Full Moment Tensors, Including Uncertainties, for Nuclear Explosions, Volcanic Events, and Earthquakes. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5099–5119. Retrieved 2020-01-15, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JB015325` doi: 10.1029/2017JB015325

Del Moral, P., Doucet, A., & Jasra, A. (2006). Sequential Monte Carlo Samplers. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *68*(3), 411–436. Retrieved 2023-08-14, from `https://www.jstor.org/stable/3879283` (Publisher: [Royal Statistical Society, Wiley])

Dreger, D., Tkalčić, H., & Johnston, M. (2000, April). Dilational Processes Accompanying Earthquakes in the Long Valley Caldera. *Science*, *288*(5463), 122–125. Retrieved 2020-08-02, from `https://www.sciencemag.org/lookup/doi/10.1126/science.288.5463.122` doi: 10.1126/science.288.5463.122

Dreger, D. S., & Helmberger, D. V. (1990, October). Broadband modelling of local earthquakes. *Bulletin of the SeismologicalSocietyof America*, *80*(5), 1162–1179.

Duputel, Z., Rivera, L., Fukahata, Y., & Kanamori, H. (2012, August). Uncertainty estimations for seismic source inversions. *Geophysical Journal International*, *190*(2), 1243–1256. Retrieved 2020-01-21, from `https://academic.oup.com/gji/article/190/2/1243/645429` doi: 10.1111/j.1365-246X.2012.05554.x

Fichtner, A., & Simutė, S. (2018). Hamiltonian Monte Carlo Inversion of Seismic Sources in Complex Media. *Journal of Geophysical Research: Solid Earth*, *123*(4), 2984–2999. Retrieved 2019-12-12, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JB015249` doi: 10.1002/2017JB015249

Ford, S. R., Dreger, D. S., & Walter, W. R. (2010, October). Network Sensitivity Solutions for Regional Moment-Tensor InversionsNetwork Sensitivity Solutions for Regional Moment-Tensor Inversions. *Bulletin of the Seismo-*

logical Society of America, *100*(5A), 1962–1970. Retrieved 2020-05-12, from http://pubs.geoscienceworld.org/bssa/article/100/5A/1962/325105/Network-Sensitivity-Solutions-for-Regional-Moment (Publisher: GeoScienceWorld) doi: 10.1785/0120090140

Goodman, J., & Weare, J. (2010, January). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1), 65–80. Retrieved 2022-05-09, from http://msp.org/camcos/2010/5-1/p04.xhtml doi: 10.2140/camcos.2010.5.65

Gradshteyn, I. S., Ryzhik, I. M., Jeffrey, A., & Zwillinger, D. (2000). *Table of Integrals, Series, and Products, Sixth Edition* (6th edition ed.). San Diego: Academic Press.

Hejrani, B., & Tkalčić, H. (2018, December). The 20 May 2016 Petermann Ranges earthquake: centroid location, magnitude and focal mechanism from full waveform modelling. *Australian Journal of Earth Sciences*, *0*(0), 1–9. Retrieved 2018-12-19, from https://doi.org/10.1080/08120099.2018.1525783 doi: 10.1080/08120099.2018.1525783

Hejrani, B., & Tkalčić, H. (2020). Resolvability of the Centroid-Moment-Tensors for Shallow Seismic Sources and Improvements From Modeling High-Frequency Waveforms. *Journal of Geophysical Research: Solid Earth*, *125*(7), e2020JB019643. Retrieved 2020-07-03, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JB019643 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020JB019643) doi: 10.1029/2020JB019643

Hejrani, B., Tkalčić, H., & Fichtner, A. (2017). Centroid moment tensor catalogue using a 3-D continental scale Earth model: Application to earthquakes in Papua New Guinea and the Solomon Islands. *Journal of Geophysical Research: Solid Earth*, *122*(7), 5517–5543. Retrieved 2019-12-27, from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JB014230 doi: 10.1002/2017JB014230

Herrmann, R. B. (2013, November). Computer Programs in Seismology: An Evolving Tool for Instruction and Research. *Seismological Research Letters*, *84*(6), 1081–1088. Retrieved 2020-02-21, from https://pubs.geoscienceworld.org/srl/article/84/6/1081/315307/Computer-Programs-in-Seismology-An-Evolving-Tool doi: 10.1785/0220110096

Hu, J., Pham, T.-S., & Tkalčić, H. (2023, January). *Point-source moment tensor inversion via a Bayesian hierarchical inversion with 2D-structure uncertainty: Implications for the 2009-2017 DPRK nuclear tests* (preprint). Preprints. Retrieved 2023-06-09, from https://essopenarchive.org/users/574509/articles/618243-point-source-moment-tensor-inversion-via-a-bayesian-hierarchical-inversion-with-2d-structure-uncertainty-implications-for-the-2009-2017-dprk-nuclear-tests?commit=649925c430c8d3866ba76bf38b09ed74bc2833eb doi: 10.22541/essoar.167397479.97900153/v1

Jost, M. L., & Herrmann, R. B. (1989, April). A Student's Guide to and Review of Moment Tensors. *Seismological Research Letters*, *60*(2), 37–57. Retrieved 2018-11-01, from https://pubs.geoscienceworld.org/ssa/srl/article/60/2/37/141825/a-student-s-guide-to-and-review-of-moment-tensors doi: 10.1785/gssrl.60.2.37

Kawakatsu, H. (1996, August). Observability of the isotropic component of a moment tensor. *Geophysical Journal International*, *126*(2), 525–544. Retrieved 2022-01-07, from https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.1996.tb05308.x doi: 10.1111/j.1365-246X.1996.tb05308.x

Kikuchi, M., & Kanamori, H. (1982, April). Inversion of complex body waves. *Bulletin of the Seismological Society of America*, *72*(2), 491–506. Retrieved 2016-

05-23, from http://www.bssaonline.org/content/72/2/491

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization.* Retrieved 2023-06-14, from https://arxiv.org/abs/1412.6980 (Publisher: arXiv Version Number: 9) doi: 10.48550/ARXIV.1412.6980

Minson, S. E., & Dreger, D. S. (2008, August). Stable inversions for complete moment tensors. *Geophysical Journal International*, *174*(2), 585–592. Retrieved 2019-01-13, from https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.2008.03797.x doi: 10.1111/j.1365-246X.2008.03797.x

Mustać, M., & Tkalčić, H. (2016). Point source moment tensor inversion through a Bayesian hierarchical model. *Geophysical Journal International*, *204*(1), 311–323. doi: 10.1093/gji/ggv458

Phạm, T.-S., & Tkalčić, H. (2021). Toward Improving Point-Source Moment-Tensor Inference by Incorporating 1D Earth Model's Uncertainty: Implications for the Long Valley Caldera Earthquakes. *Journal of Geophysical Research: Solid Earth*, *126*(11), 2021JB022477. Retrieved 2021-11-08, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2021JB022477 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021JB022477) doi: 10.1029/2021JB022477

Roberts, G. O., & Rosenthal, J. S. (2001). Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science*, *16*(4), 351–367. Retrieved 2023-08-14, from https://www.jstor.org/stable/3182776 (Publisher: Institute of Mathematical Statistics)

Sambridge, M., Jackson, A., & Valentine, A. P. (2022, April). Geophysical inversion and optimal transport. *Geophysical Journal International*, ggac151. Retrieved 2022-04-26, from https://academic.oup.com/gji/advance-article/doi/10.1093/gji/ggac151/6572363 doi: 10.1093/gji/ggac151

Silwal, V., & Tape, C. (2016). Seismic moment tensors and estimated uncertainties in southern Alaska. *Journal of Geophysical Research: Solid Earth*, *121*(4), 2772–2797. Retrieved 2022-01-18, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2015JB012588 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015JB012588) doi: 10.1002/2015JB012588

Stein, S., & Wysession, M. (2003). *An Introduction to Seismology, Earthquakes, and Earth Structure* (1edition ed.). Malden, MA: Blackwell Publishing.

Stähler, S. C., & Sigloch, K. (2014, November). Fully probabilistic seismic source inversion - Part 1: Efficient parameterisation. *Solid Earth*, *5*(2), 1055–1069. Retrieved 2020-01-15, from https://www.solid-earth.net/5/1055/2014/ doi: https://doi.org/10.5194/se-5-1055-2014

Tape, W., & Tape, C. (2012). A geometric setting for moment tensors. *Geophysical Journal International*, *190*(1), 476–498. doi: 10.1111/j.1365-246X.2012.05491.x

Tape, W., & Tape, C. (2015, September). A uniform parametrization of moment tensors. *Geophysical Journal International*, *202*(3), 2074–2081. Retrieved 2020-03-23, from https://academic.oup.com/gji/article/202/3/2074/613765 (Publisher: Oxford Academic) doi: 10.1093/gji/ggv262

Thurin, J., Tape, C., & Modrak, R. (2022, October). Multi-Event Explosive Seismic Source for the 2022 Mw 6.3 Hunga Tonga Submarine Volcanic Eruption. *The Seismic Record*, *2*(4), 217–226. Retrieved 2023-06-27, from https://doi.org/10.1785/0320220027 doi: 10.1785/0320220027

Vackář, J., Burjánek, J., Gallovič, F., Zahradník, J., & Clinton, J. (2017, August). Bayesian ISOLA: new tool for automated centroid moment tensor inversion. *Geophysical Journal International*, *210*(2), 693–705. Retrieved 2019-12-29, from https://academic.oup.com/gji/article/210/2/693/3747443 doi: 10.1093/gji/ggx158

Vasyura-Bathke, H., Dettmer, J., Dutta, R., Mai, P. M., & Jónsson, S. (2021,

March). Accounting for theory errors with empirical Bayesian noise models in nonlinear centroid moment tensor estimation. *Geophysical Journal International*, *225*(2), 1412–1431. Retrieved 2021-03-21, from https://doi.org/10.1093/gji/ggab034 doi: 10.1093/gji/ggab034

Viltres, R., Nobile, A., Vasyura-Bathke, H., Trippanera, D., Xu, W., & Jónsson, S. (2022, March). Transtensional Rupture within a Diffuse Plate Boundary Zone during the 2020 Mw 6.4 Puerto Rico Earthquake. *Seismological Research Letters*, *93*(2A), 567–583. Retrieved 2023-06-19, from https://pubs.geoscienceworld.org/srl/article/93/2A/567/610483/Transtensional-Rupture-within-a-Diffuse-Plate doi: 10.1785/0220210261

Zhao, L.-S., & Helmberger, D. V. (1994, February). Source estimation from broadband regional seismograms. *Bulletin of the Seismological Society of America*, *84*(1), 91–104. Retrieved 2023-06-19, from https://doi.org/10.1785/BSSA0840010091 doi: 10.1785/BSSA0840010091

Zhu, L., & Ben-Zion, Y. (2013, August). Parametrization of general seismic potency and moment tensors for source inversion of seismic waveform data. *Geophysical Journal International*, *194*(2), 839–843. Retrieved 2023-06-19, from https://academic.oup.com/gji/article/194/2/839/2111069 doi: 10.1093/gji/ggt137

Zhu, L., & Helmberger, D. V. (1996, October). Advancement in source estimation techniques using broadband regional seismograms. *Bulletin of the Seismological Society of America*, *86*(5), 1634–1641. Retrieved 2023-06-14, from https://pubs.geoscienceworld.org/bssa/article/86/5/1634/120218/Advancement-in-source-estimation-techniques-using doi: 10.1785/BSSA0860051634

Figure 1.

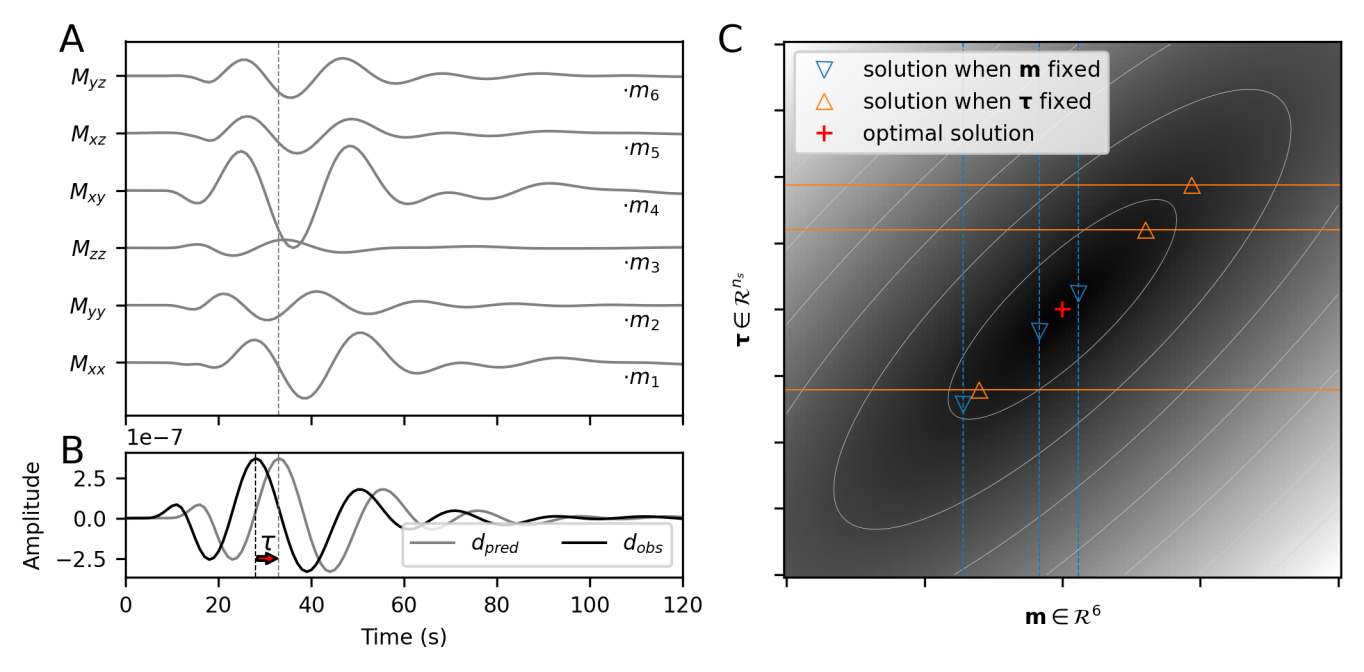

A
M_{yz}
M_{xz}
M_{xy}
M_{zz}
M_{yy}
M_{xx}
$\cdot m_6$
$\cdot m_5$
$\cdot m_4$
$\cdot m_3$
$\cdot m_2$
$\cdot m_1$
B
1e−7
Amplitude
2.5
0.0
−2.5
τ
d_{pred}
d_{obs}
0
20
40
60
80
100
120
Time (s)
C
solution when **m** fixed
solution when **τ** fixed
optimal solution
$\boldsymbol{\tau} \in \mathcal{R}^{n_s}$
$\mathbf{m} \in \mathcal{R}^6$

Figure 2.

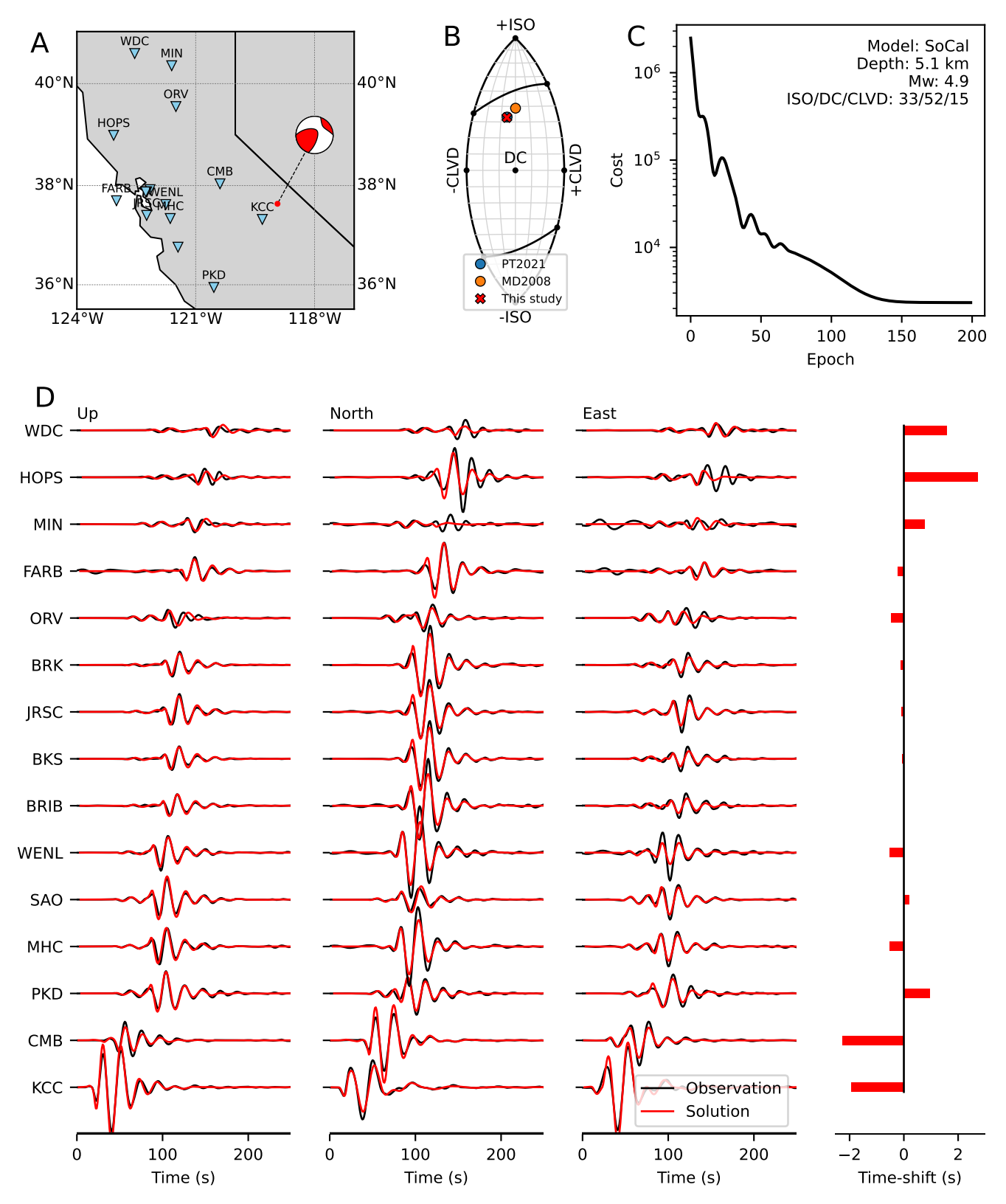

A
WDC
MIN
ORV
HOPS
CMB
FARB
WENL
JRSC
MHC
KCC
PKD
40°N
38°N
36°N
124°W
121°W
118°W
B
+ISO
-ISO
-CLVD
+CLVD
DC
PT2021
MD2008
This study
C
Model: SoCal
Depth: 5.1 km
Mw: 4.9
ISO/DC/CLVD: 33/52/15
Cost
Epoch
D
Up
North
East
WDC
HOPS
MIN
FARB
ORV
BRK
JRSC
BKS
BRIB
WENL
SAO
MHC
PKD
CMB
KCC
Observation
Solution
Time (s)
Time-shift (s)

Figure 3.

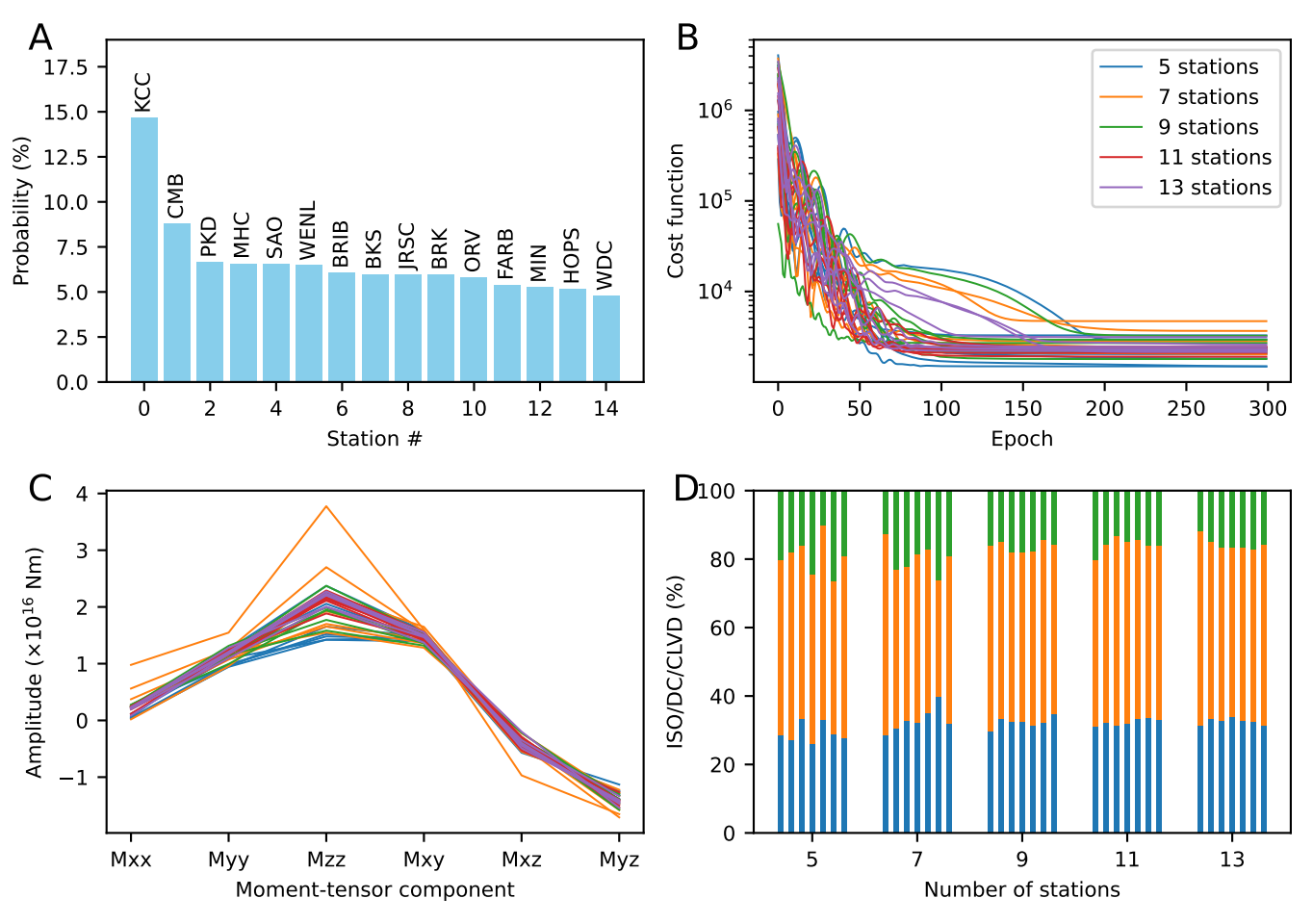

A
Probability (%)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
KCC
CMB
PKD
MHC
SAO
WENL
BRIB
BKS
JRSC
BRK
ORV
FARB
MIN
HOPS
WDC
0
2
4
6
8
10
12
14
Station #
B
Cost function
10^6
10^5
10^4
5 stations
7 stations
9 stations
11 stations
13 stations
0
50
100
150
200
250
300
Epoch
C
Amplitude ($\times 10^{16}$ Nm)
4
3
2
1
0
−1
Mxx
Myy
Mzz
Mxy
Mxz
Myz
Moment-tensor component
D
ISO/DC/CLVD (%)
100
80
60
40
20
0
5
7
9
11
13
Number of stations

Figure 4.

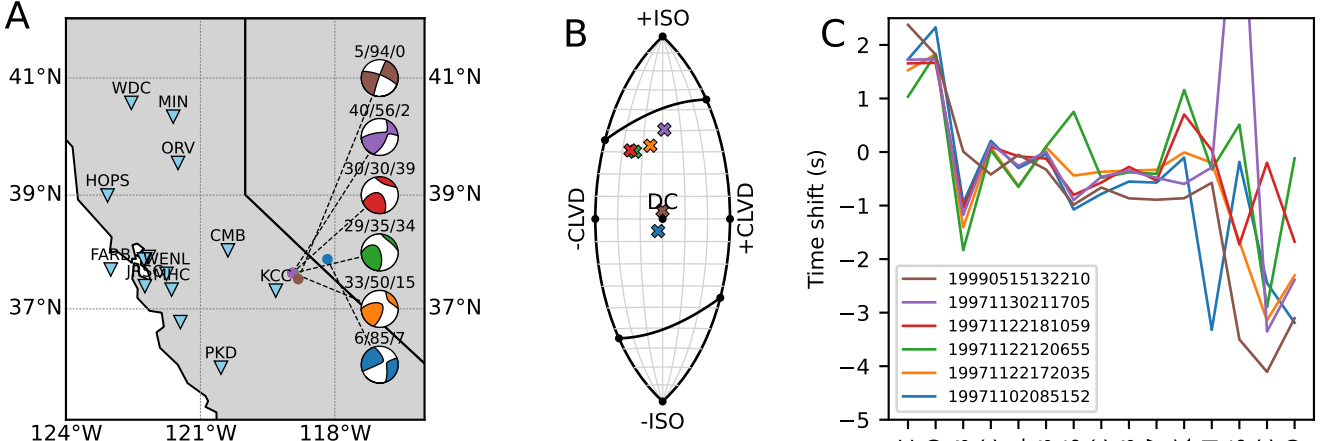
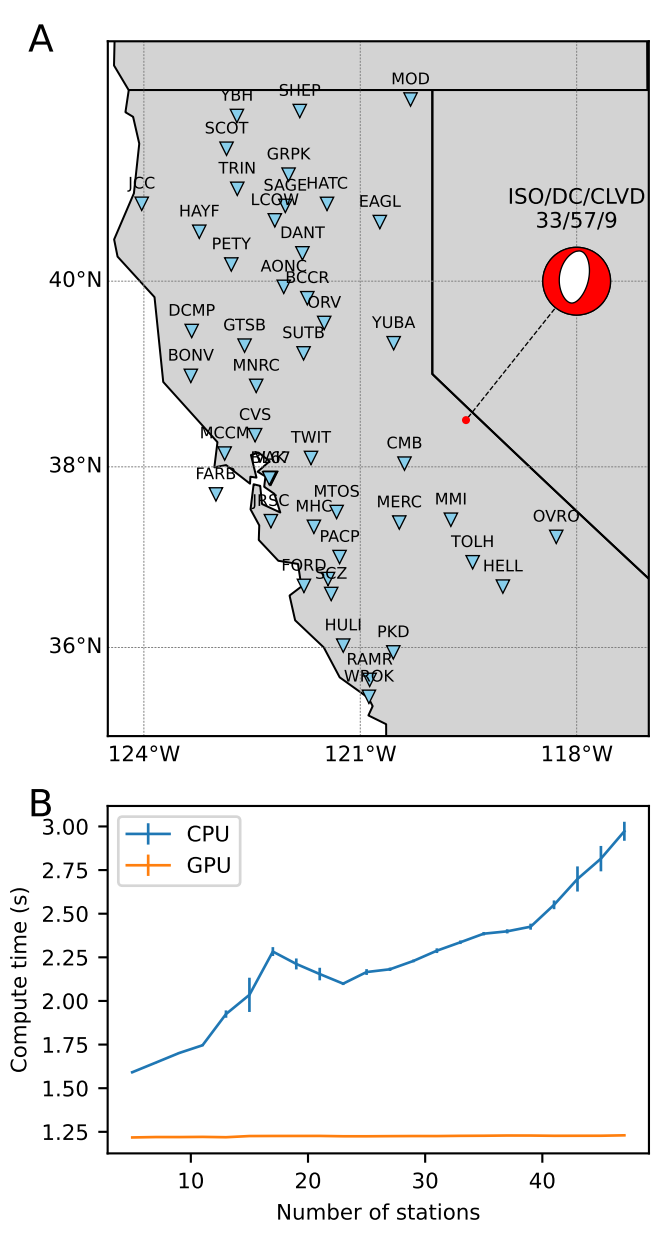

A
41°N
39°N
37°N
124°W
121°W
118°W
WDC
MIN
ORV
HOPS
CMB
FARB
WENL
JRSC
MHC
KCC
PKD
5/94/0
40/56/2
30/30/39
29/35/34
33/50/15
6/85/7
B
+ISO
-ISO
-CLVD
+CLVD
DC
C
Time shift (s)
2
1
0
−1
−2
−3
−4
−5
19990515132210
19971130211705
19971122181059
19971122120655
19971122172035
19971102085152
KCC
SAO
BRIB
MHC
WENL
CMB
BKS
JRSC
FARB
ORV
BRK
MIN
HOPS
WDC
PKD

Figure 5.


A
MOD
YBH
SHEP
SCOT
GRPK
TRIN
JCC
SAGE
HATC
LCCA
EAGL
HAYF
ISO/DC/CLVD
33/57/9
DANT
PETY
AONC
BCCR
40°N
ORV
DCMP
GTSB
SUTB
YUBA
BONV
MNRC
CVS
MCCM
TWIT
CMB
38°N
FARB
MTOS
MHC
MERC
MMI
OVRO
PACP
TOLH
FORD
SGZ
HELL
HULI
PKD
RAMR
36°N
124°W
121°W
118°W
B
CPU
GPU
Compute time (s)
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
10
20
30
40
Number of stations