

JGR Solid Earth

RESEARCH ARTICLE

10.1029/2023JB027334

Key Points:

- A neural network is used to double the number of earthquakes for source analyses compared to previous studies by improving the data quality
- Denoising teleseismic waves improves the source signature in Mw5.0–6.0 events and reduces uncertainties over all magnitudes
- Large deep earthquake ruptures are dissipative and compact relative to crustal earthquakes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Q. Shi,
qibins@uw.edu

Citation:

Shi, Q., & Denolle, M. A. (2023). Improved observations of deep earthquake ruptures using machine learning. *Journal of Geophysical Research: Solid Earth*, 128, e2023JB027334. <https://doi.org/10.1029/2023JB027334>

Received 23 JUN 2023

Accepted 2 DEC 2023

Author Contributions:

Conceptualization: Qibin Shi, Marine A. Denolle

Data curation: Qibin Shi

Formal analysis: Qibin Shi

Funding acquisition: Marine A. Denolle

Investigation: Qibin Shi, Marine A. Denolle

Methodology: Qibin Shi, Marine A. Denolle

Project Administration: Qibin Shi, Marine A. Denolle

Resources: Marine A. Denolle

Software: Qibin Shi

Supervision: Marine A. Denolle

Validation: Qibin Shi

Visualization: Qibin Shi

Writing – original draft: Qibin Shi

Writing – review & editing: Qibin Shi, Marine A. Denolle

Improved Observations of Deep Earthquake Ruptures Using Machine Learning

Qibin Shi¹  and Marine A. Denolle¹ 

¹Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA

Abstract Elevated seismic noise for moderate-size earthquakes recorded at teleseismic distances has limited our ability to see their complexity. We develop a machine-learning-based algorithm to separate noise and earthquake signals that overlap in frequency. The multi-task encoder-decoder model is built around a kernel pre-trained on local (e.g., short distances) earthquake data (Yin et al., 2022, <https://doi.org/10.1093/gji/ggac290>) and is modified by continued learning with high-quality teleseismic data. We denoise teleseismic P waves of deep Mw5.0+ earthquakes and use the clean P waves to estimate source characteristics with reduced uncertainties of these understudied earthquakes. We find a scaling of moment and duration to be $M_0 \propto \tau^4$, and a resulting strong scaling of stress drop and radiated energy with magnitude ($\Delta\sigma \propto M_0^{0.21}$ and $E_R \propto M_0^{1.24}$). The median radiation efficiency is 5%, a low value compared to crustal earthquakes. Overall, we show that deep earthquakes have weak rupture directivity and few subevents, suggesting a simple model of a circular crack with radial rupture propagation is appropriate. When accounting for their respective scaling with earthquake size, we find no systematic depth variations of duration, stress drop, or radiated energy within the 100–700 km depth range. Our study supports the findings of Poli and Prieto (2016, <https://doi.org/10.1002/2016jb013521>) with a doubled amount of earthquakes investigated and with earthquakes of lower magnitudes.

Plain Language Summary The vibration of the Earth's ground recorded at seismometers carries the seismic signatures of distant earthquakes superimposed to the Earth's natural or anthropogenic noise surrounding the seismic station. We use artificial intelligence technology to separate the weak signals of distant earthquakes from other sources of ground vibrations unrelated to the earthquakes. The separated signal provides new insights into earthquakes, especially those within the Earth's deep interior, most of which have not been investigated due to noise levels. In contrast with shallow earthquakes, deep earthquakes are less efficient at radiating energy, though they exhibit a higher rate of increase in both stress drop and radiated energy as they grow. This may suggest that deep earthquakes tend to be more confined fault surfaces. A dual mechanism between nucleation in the subduction-zone core and propagation of larger events in the dry mantle explains our observations.

1. Introduction

Deep earthquakes are understudied because they tend not to generate shaking-induced damage, only rarely generate surface displacement (Luo et al., 2023; Park et al., 2023; Steblov et al., 2014), and their extreme remoteness yields poor seismic signals on surface sensors. They usually occur in a region that surrounds the deep portion of the subducted oceanic lithosphere. The mechanisms that lead to the unstable seismic slip of deep earthquakes are still debated (Zhan, 2020). Indeed, the rheology of Earth materials does not favor brittle failure below about 70 km, thus requiring mechanisms different from shallow earthquakes. A minimum seismicity rate is reached at a depth of about 300 km (Frohlich, 1989; Green & Houston, 1995; Kirby et al., 1996; Zhan, 2020), indicating different mechanisms operate the intermediate (above 300 km) and deep-focus earthquakes (below 300 km). Previous studies have revealed somewhat complicated characteristics of the deep earthquakes (Knopoff & Randall, 1970; Ye et al., 2016). The focal mechanisms of deep earthquakes usually show non-double-couple components (Knopoff & Randall, 1970), implying more complex rupture processes than simple shear dislocation on faults with uniform fault geometries. The non-double-couple moment tensor could also be partially attributed to the anisotropic features of the slab rock fabric (Li et al., 2018). The greater stress drop claimed for deep earthquakes is well explained by the increased rigidity with depth (Vallée, 2013), though the low-stress drop is occasionally found for deep earthquakes (Turner et al., 2022). Multiple investigations found a strong magnitude dependence of the stress drop, which may be interpreted as dynamic weakening mechanisms (Oth et al., 2009; Poli & Prieto, 2016; Prieto et al., 2013; Radulian & Popa, 1996). Deep earthquakes follow Gutenberg-Richter

law (Gutenberg & Richter, 1949) but have depleted aftershock productivity compared to shallow earthquakes (Dascher-Cousineau et al., 2020; Ye et al., 2020).

The presence of deep earthquakes within the subducted slab provides an interesting window to explore the physical processes of subduction. Zhan (2020) reviewed the three leading mechanisms that favor dynamic rupture of deep earthquakes: (a) mineral dehydration from metamorphic processes that release fluids and lubricate faults (i.e., dehydration embrittlement), (b) phase transformation that changes mineral density and volume, and (c) thermal runaway that lowers fault friction from shear heating. The fluids released by mineral dehydration are thought to explain the double-seismic zone (DSZ) (Abers et al., 2013; Brudzinski et al., 2007; Hacker et al., 2003; Yamasaki & Seno, 2003). Whether the released water can penetrate the slab core (Boneh et al., 2019; Green & Houston, 1995) and be transported deeper in the mantle is still under debate (Pearson et al., 2014; Plümpner et al., 2017; Schmandt et al., 2014; Sobolev et al., 2019; Tschauner et al., 2018).

Teleseismic observations of deep earthquakes are the most common data available to study these earthquakes. Although regional observations of deep earthquakes may be available at some area with densely distributed seismic networks (Chu et al., 2019; Kita & Katsumata, 2015; Ko & Kuo, 2016; Nishitsuji & Mori, 2014) or with temporary seismic arrays (Tian et al., 2022), teleseismic observations are usually the most important data for the most regions in the world. Because small events are more frequent than large earthquakes, moderate-size earthquakes (Mw5.0–6.0) could provide crucial constraints on the rupture mechanisms of deep earthquakes. However, elevated seismic noise has limited our ability to investigate the dynamics of moderate-size earthquakes (Mw5.0–6.0) from teleseismic distances. The source analyses of deep earthquakes have been conducted with only the high signal-to-noise ratio (SNR) data of Mw5.8+ earthquakes (Poli & Prieto, 2014, 2016), leaving a vast number of moderate-magnitude earthquakes ignored given then with lower SNR waveforms. Furthermore, SNR-based data selection of teleseismic P waves may result in azimuthal biases with azimuths and take-off angles due to the radiation pattern. The intrinsic variability of rupture mechanisms associated with magnitudes has not been fully explored. Hence, the events with a wider range of SNR may be more meaningful.

The superposition of seismic noise and signal at overlapping frequencies poses challenges to the traditional Fourier-based noise removal approaches (Douglas, 1997). Other time-frequency methods are useful in separating the overlapped spectra but require extensive human intervention (Chang et al., 2000; Donoho & Johnstone, 1994; Mousavi & Langston, 2017; Stockwell et al., 1996). The recent development of deep neural networks for seismological research has repeatedly demonstrated its potential for extracting coherent earthquake features from noisy seismic observations. Several recent studies have applied machine learning to denoise the signals in the time-frequency domain with the assumption that local earthquake and noise signals have distinct Fourier spectra. Zhu et al. (2019) converted seismic time series (seismograms) of local earthquakes to a time-frequency representation and developed a deep convolutional neural network to extract the earthquake signals in a time-frequency latent space. Using direct time-frequency input has gained traction (Chen et al., 2022; Saad & Chen, 2020). In fact, the time-frequency information may also be utilized implicitly by appropriate convolutional layers considered multi-frequency-band “filters” in the time domain. Using that concept, Novoselov et al. (2022) showed that recurrent neural networks could separate overlapping seismic signals produced by distinct sources. Yin et al. (2022) combined two-branch encoder-decoder and recurrent neural networks to compose the WaveDecompNet, which has been proven effective in reconstructing local earthquake and noise waveforms. Yin et al. (2022) demonstrated that even the clean noise waveforms improved the coherence of noise single-station cross-correlations for ambient noise seismology.

Challenges remain in using these existing models to denoise teleseismic recordings. First, teleseismic waveforms have a much lower SNR than local or regional waveforms for the same earthquake magnitude, mainly due to the geometrical spreading and attenuation. Second, the attenuation of global seismic phases distorts the signal such that signal frequencies overlap with the microseismic signals in velocity seismograms.

This study uses a multi-task encoder-decoder to denoise the teleseismic waves of global M5.0+ earthquakes, a method that we name “DenoTe” (Shi, 2023). The neural network takes the architecture of WaveDecompNet (Yin et al., 2022) as a kernel to extract high-level features of the teleseismic body waves and uses convolutional layers to reconstruct the denoised signals and pure noise signals. We add a layer on the top and bottom of the kernel network to adjust the input window lengths. Our training data comprises teleseismic data from the International Federation of Digital Seismograph Networks (FDSN) and Global Seismographic Network (GSN) for Mw5.0+ earthquakes of the 2000–2021 International Seismological Centre (ISC) earthquake catalog (International

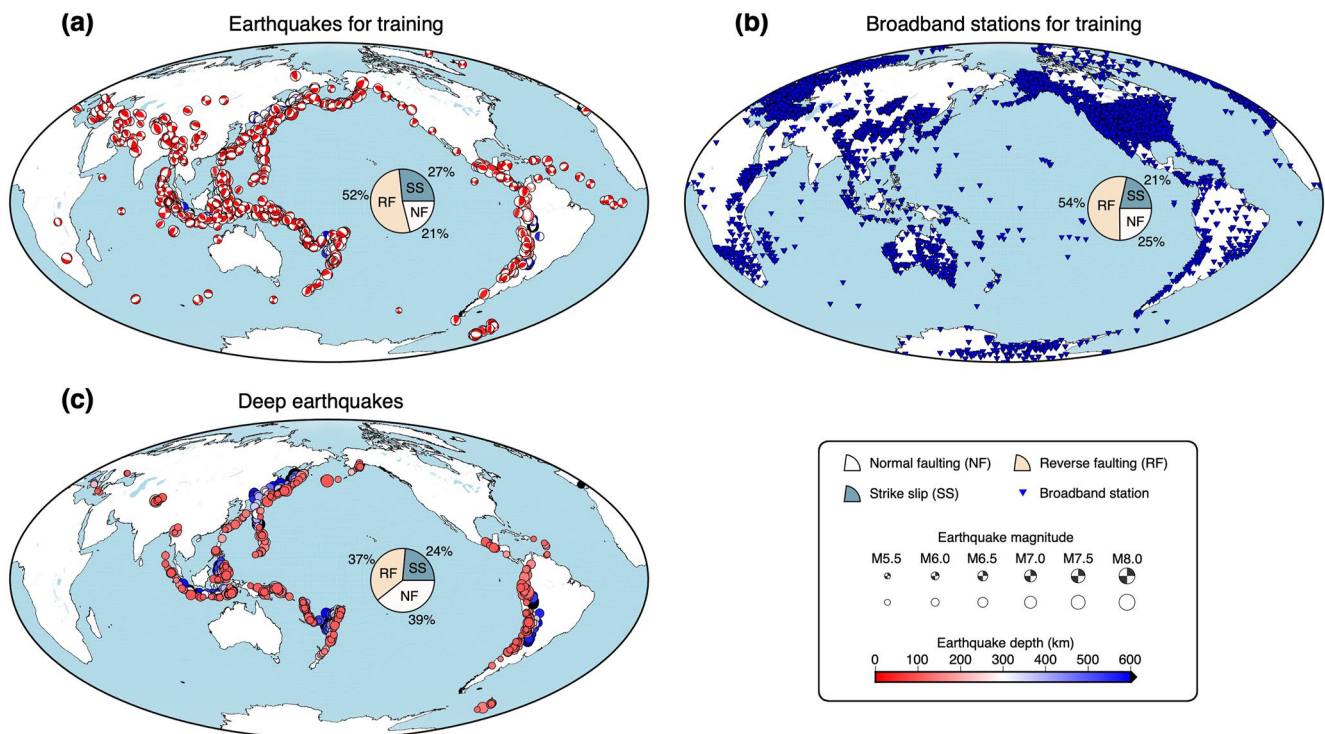


Figure 1. Earthquakes and seismic stations. (a) The 1148 earthquakes with high-SNR recordings were used as training data. (b) The FDSN and GSN broadband stations recorded the 45,262 high-SNR teleseismic waveforms of the 1148 earthquakes. (c) The 920 deep earthquakes with low-SNR teleseismic waveforms labeled with focal mechanisms are denoised and tested in this study. The legends and color scale are indicated on the lower-right. We use the depth range of 0–600 km for the better visualization of the color scale and note that the actual data for the depth range of 0–100 km are not presented using this color scale throughout this article.

Seismological Centre, 2022). The pre-trained kernel is updated through transfer learning. We denoise the teleseismic body waves to extract P-wave pulses of deep Mw5.0+ earthquakes. We estimate several source parameters: pulse duration and rupture directivity using relative duration measurements and radiated energy, stress drop, and fracture energy using denoised P-wave spectra. We discuss the strong scaling of these properties with earthquake magnitude in contrast with the typical scaling of crustal earthquakes and the possible dual mechanisms that explain intermediate and deep earthquakes.

2. Data Preparation

We use supervised learning to separate the earthquake and noise waveforms from their combined form. The amount, diversity, and accuracy of the training data greatly impact learning performance. The volume of high-quality earthquake records from global seismic networks has grown vastly in the past two decades. We extract 1148 Mw5.0+ earthquakes from the 2000–2021 ISC earthquake catalog (International Seismological Centre, 2022) based on focal mechanisms to ensure a relatively even number of strike-slip (306), normal-faulting (242), and reverse-faulting (600) earthquake types. Specifically, we classify the event mechanisms using the rake: reverse faults are between 45° and 135°, normal faults are between –45° and –135°, and the rests are strike-slip events. The extracted earthquake list includes events from diverse seismic regions and depths ranging from the surface to 700 km (Figure 1a).

To prepare the labels of “clean” P waves seismic waveforms, we download data from all broadband seismometers available from the FDSN stations selected at teleseismic angular distances between 30° and 90° to avoid Moho and core reflected and converted phases. The P waves of Mw5.0–5.9 are noisy in general and, thus, tend not to be included in the training data given our signal-to-noise ratio-based selection criteria. We calculate the P-wave arrival time based on the catalog origin time and hypocentral location using an Obspy implementation of Tau-P (Beyreuther et al., 2010; Crotwell et al., 1999) in an IASPI91 Earth model (Kennet, 1991). We then downsample the three-component ground velocity waveforms down to 10 Hz and cut a wide time window starting from 2,500 s

before and 2,500 s after the P-arrival. Keeping wide time windows allows sufficient data before and after the P for our data augmentation strategy discussed below. We then calculate the amplitude-based SNR using a noise window (75–10 s before) and a signal window (0–75 s after the P-wave arrival) with the following definition,

$$SNR = \frac{A_S}{A_N}, \quad (1)$$

where A_S and A_N are the standard deviations of the amplitudes of the signal window and noise window, respectively. We only select the clean P-wave labels with SNR higher than 25 for training. We gathered 45,262 high-SNR P waves of 1,148 earthquakes of magnitude Mw5.5+. To generate realistic noise waveforms, we extract a 150-s noise window before each P wave arrival time and consider it as the noise signal specific to the station. Our data selection provides 45,262 earthquake traces and 45,262 noise traces, each composed of three-component seismograms. The proportions of waveforms generated by the strike-slip, normal-faulting, and reverse-faulting events are 21%, 25%, and 54%, respectively (Figure 1b).

3. Denoising

We develop, train, and apply a multi-task encoder-decoder to denoise the teleseismic P waves in the time domain. We adapt from an existing model architecture by Yin et al. (2022) to use teleseismic data.

3.1. Neural Network Architecture

We expand from the encoder-decoder network of Yin et al. (2022) to adapt to longer input window lengths. We follow a similar style as WaveDecompNet in Yin et al. (2022). Because the teleseismic waveforms have distinct low-level features from the local waveforms, we stack the WaveDecompNet kernel with feature extraction layers. The stacked neural network on the top encoder branch is a two-layer convolutional neural network (CNN) with a one-layer fully connected layer (FCNN) on the optimal training performance. Next, we introduce the architecture of the two-branch encoder-decoder (Figure 2) and the strategy to enhance training efficiency.

Similar to Yin et al. (2022), we use a stride of two after each CNN layer to avoid aliasing (Zhang, 2019). A skip connection is introduced after the first CNN layer to retain the fine scale of the feature. Compared to the single-branch prediction of either the earthquake or noise signal (Novoselov et al., 2022; Zhu et al., 2019), our multi-task model (i.e., two-branch prediction) depends on the efficiency of feature extraction for both earthquake and noise signals.

The data is normalized using standard scaling (removing the mean and normalizing by the data standard deviation) and can be rescaled after the wavefield separation by the same scaling factor. In the following analysis, where we measure duration estimates and normalize the data to seismic moment, we do not rescale the data after denoising.

3.2. Data Augmentation

Training the model with 60% of the overall data is insufficient to yield a satisfying model performance (see details below). Therefore, we proceed with a data augmentation approach to improve model training. We conduct a three-step data augmentation to increase the diversity of the training data (Figure 3), which is most important to the generalization of neural networks. The training data is more likely selected from higher magnitude earthquakes (i.e., Mw6.0+), which tend to have longer source duration and thus tend to generate relatively lower-frequency signals compared to the more frequent smaller earthquakes. Hence, the raw training data lacks high-frequency information, such as those expected for lower-magnitude earthquakes (Mw5.0–6.0). To generate high-frequency data compatible with these small earthquakes, we augment the training data of earthquake waveforms by squeezing the seismogram along the time axis. The squeezing ratio is randomly sampled from 1,2,...8 with equal probability (i.e., 12.5% for all ratios). We then shift waveforms to avoid the case of the denoising algorithm memorizing the stationary P-wave arrival time (Zhu et al., 2020). We take the theoretical P arrival time as the original zero and then shift waveforms using a uniform probability between ± 75 s. After shifting, we trim the time series to the -75 s $\sim$ $+75$ s time window. Thus, the trimmed waveforms mostly include the P wave onsets. In the final augmentation step, we stack each 150-s trace with the 150-s amplified noise extracted from pre-P noise at the

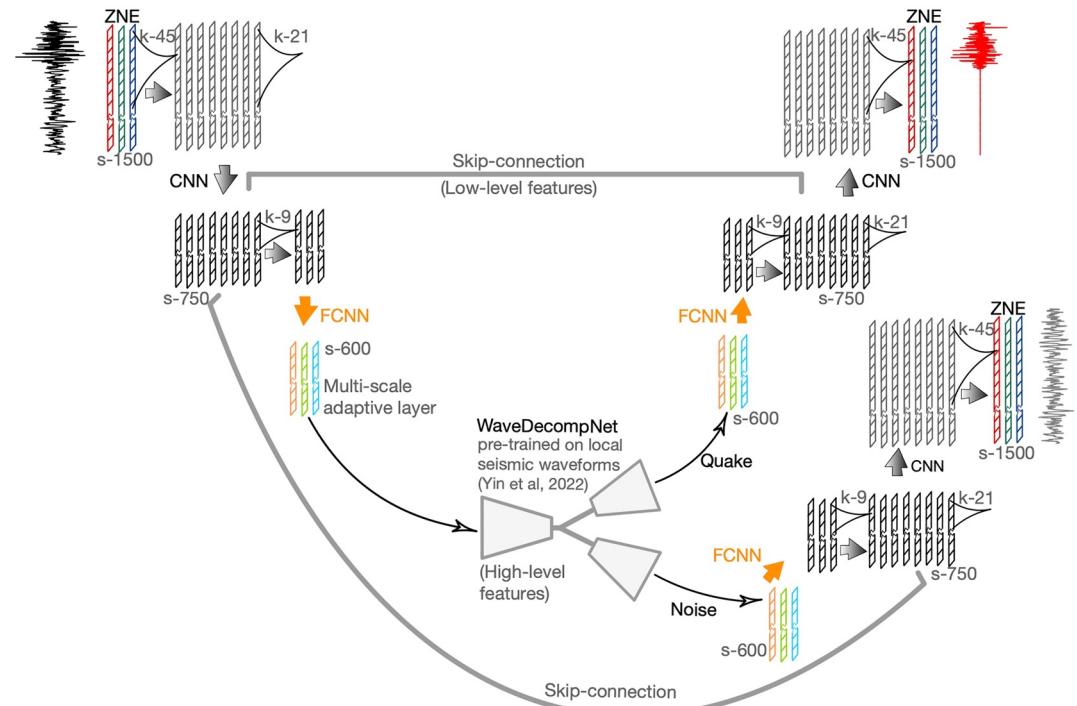


Figure 2. Architecture of the teleseismic wave denoiser, DenoTe. DenoTe is constructed based on the U-net with symmetric structures in the encoding and decoding branches of WaveDecompNet (Yin et al., 2022). The neural network reads composite earthquake waveforms (black) and predicts earthquake (red) and noise (gray) signals through the two output branches, which have the same structure and length. The sizes of feature maps in each layer (s-1500, s-750, and s-600) and kernel length (k-9, k-21, and k-45) of each convolutional filter are indicated as gray text. The number of traces represents the number of channels. CNN: convolutional neural network. FCNN: fully connected neural network.

same channel. A random SNR (as defined in Equation 1) between 0.5 and 10 is selected to give earthquake and noise relative weights in the combined, “noisy” waveform. The three-step augmentation—stretching, shifting, and adding noise—is performed repeatedly in every training epoch with randomly selected parameters. The diversity

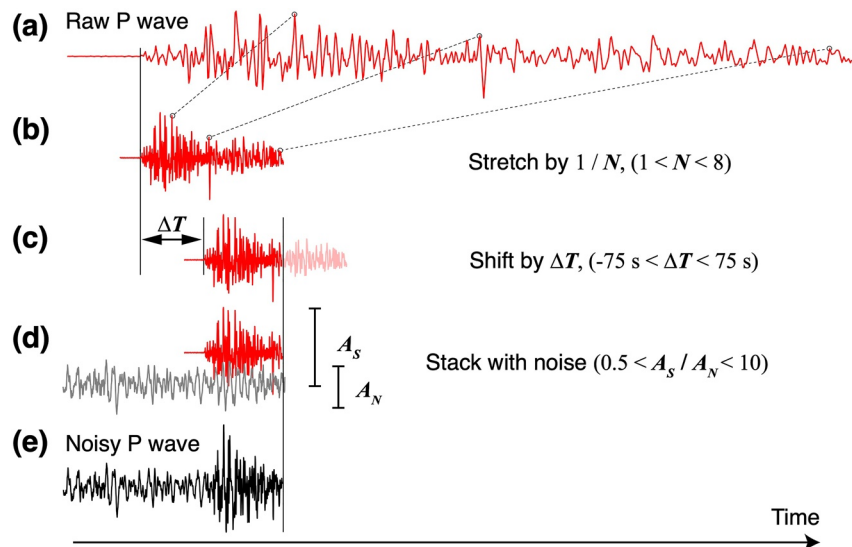


Figure 3. The three steps of data augmentation. (a) The raw high-SNR P wave (red) is (b) stretched, (c) shifted along the time axis, and (d) vertically scaled, and (e) stacked with the noise (gray) extracted from the same station to compose the noisy waveform (black).

of the data is enhanced with each additional training step (epoch), which reduces the possibility of overfitting the training data (Zhu et al., 2020).

3.3. Training

We train DenoTe using the composed waveform data and high-quality labels of the P-wave and noise signals. We first shuffle and then split the entire data set and corresponding labels into three subsets: 60% for training, 20% for validation, and 20% for testing. Data augmentation (Section 3.2) is done after the split, ensuring no data exchange among subsets or no data leakage leading to unrealistic testing scores. The validation and test data are also augmented data sets after data augmentation of the original data. Training is greatly improved thanks to data augmentation.

The main criterion for proper denoising is the similarity between the predicted and labeled waveforms for both earthquake and noise time series. To improve from the classic loss function mean-squared error (MSE) and focus on wiggle-by-wiggle reconstruction, we define a new loss function that combines the Pearson correlation coefficient (CC) and the MSE of the residual waveforms: $\text{loss} = \text{MSE} + (1 - \text{CC})$. The CC is independent of the absolute wave amplitude, typically between -1.0 and 1.0 , such that $(1 - \text{CC})$ varies between 0 and 2 . In comparison, the MSE typically ranges between 0 and 1 . Different weighting choices are tested between MSE and $(1 - \text{CC})$. We find by trial and error an equal weighting between both is optimal for reducing the waveform misfit.

We train for up to 200 epochs and set up an early stopping mechanism when the minimum validation loss is not updated for 20 consecutive epochs. We randomly divide the training subset into 177 mini-batches containing 256 three-component waveforms. The learning rate is fixed at 0.001 , combined with an adaptive momentum (ADAM) to control the step size in the gradient-descent process. This training process is efficient and converges at a low loss of about 0.45 after 140 epochs (see Figure S1 in Supporting Information S1). The validation loss computed for every epoch closely follows the training loss. The final testing loss is 0.45 (Figure S1 in Supporting Information S1), similar to the training and validation losses. The training, validation, and test losses suggest that the neural network does not over-fit the training data and may generalize to diverse teleseismic waves. In Figure 4, we compare the ground truth waveform and the predicted waveforms (P wave and noise), both matching well the amplitude of the pulse and the phases in the direct and coda waves of P and S waves. In Figure S13 of the Supporting Information S1, we plot the statistics of the accuracy scores of 2000 testing waveforms, including the explained variance (EV) and cross-correlation coefficient (CC) of both the earthquake and noise branches. The high EV for even the lowest-SNR data as an input indicates that our denoiser is well-trained and generalizable. Note that CC correlates with the input SNR, although generally above 0.7 , implying CC is a more strict criterion to enhance DenoTe's learning efficiency.

3.4. Predicting (Denoising) the P Waves

We apply DenoTe to 3,079 Mw5.0+ deep earthquakes between 1/1/2000 and 12/31/2021, of which 920 are labeled with focal mechanisms (217 strike-slip, 341 reverse-faulting and 362 normal faulting events as shown in Figure 1c). The data is normalized before prediction and rescaled after wavefield separation using standard scaling.

For subsequent validation of the source characteristics, we select the raw, noisy P waves with $\text{SNR} > 2$ (as defined in Equation 1) and extract the denoised P waves through DenoTe. This ensures that the post-processing analysis is only selecting data that could have been included in previous analysis and should limit the effect of artifacts generated by the model (though these were minimal when using the WaveDecompNet kernel Yin et al. (2022)).

The first-order source processes are better analyzed from displacement waveforms since these are proportional to the moment-rate function in the far-field seismograms. Therefore, we integrate all denoised velocity waveforms to displacement and normalize them to their maximum absolute amplitude. We show waveform examples from two earthquakes, original and denoised waveforms, sorted by station azimuth relative to the earthquake epicenter, aligned using cross-correlation (Figure 5). We find a systematic improvement of the P wave signal-to-noise ratio for a broad range of frequencies after denoising.

We find, in general, that the noise is considerably reduced: pre-P signals have much lower amplitudes and low-frequency noises after the P and are also absent in the post-P pulse. Because of the noise removal, it is a lot

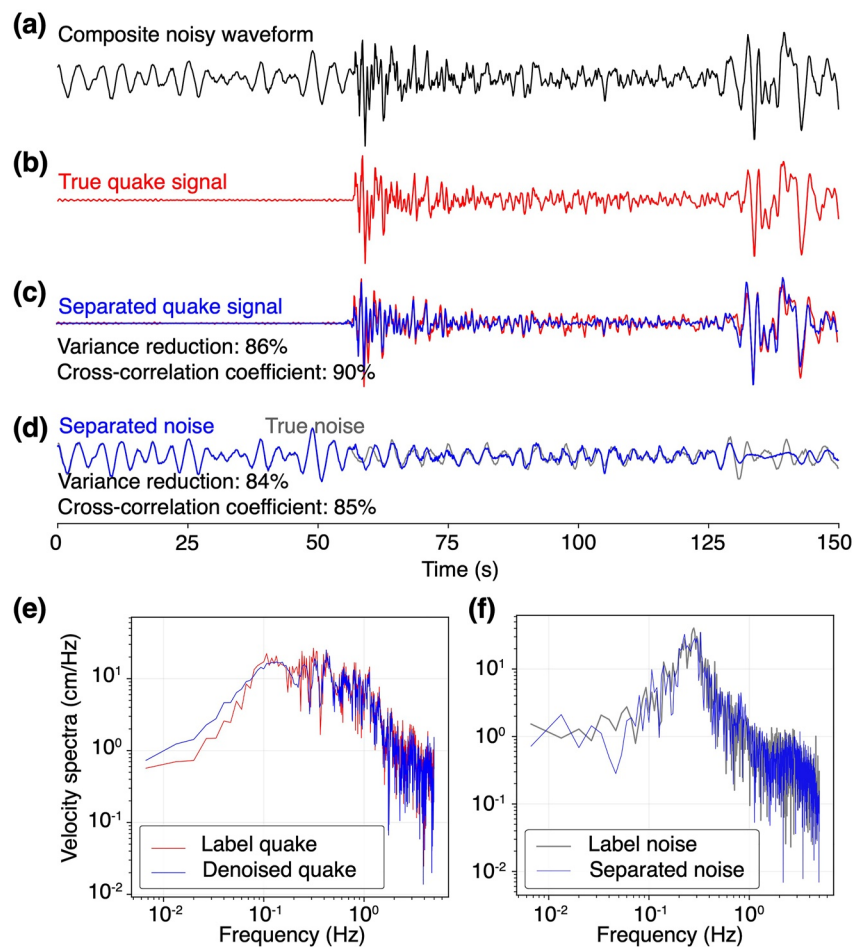


Figure 4. Example of DenoTe's performance. In the time domain: (a) composite waveform, (b) (label) earthquake signal (label data, P-wave, its coda, and the direct S wave), (c) comparison between the labeled (red) and predicted (blue) earthquake signals (and their variance reduction and correlation coefficient) (d) comparison between the labeled (red) and predicted (blue) noise signals (and their variance reduction and correlation coefficient). In the frequency domain: (e) comparison between the velocity spectra of the label and predicted earthquake data and (f) comparison between the velocity spectra of the label and predicted noise data.

easier to visualize and automatically measure pulse width. The statistical improvement of SNR for the deep-focus and intermediate-depth earthquakes is shown in Figure 5f.

4. Source Parameters

The goal of this study is to improve the quality of the source parameters of the deep Mw5.0+ earthquakes. Source parameters are directly extracted from the time domain (source duration and directivity) or measured from the spectral domain after correction for the path effects (corner frequency, stress drop, radiated energy, and radiation efficiency).

There exist many sources of uncertainty when estimating earthquake parameters: seismogram noise, distribution of the stations for a full sampling of the focal sphere, attenuation effects on the seismograms, limited frequency bandwidth of the data, simplified model, etc. Recent efforts by the community have addressed some of these issues. Bindi et al. (2023) propagated errors from data truncation, model assumptions, and choices in site amplification correction to derive uncertainties and assume normal distributions. Abercrombie (2021) illustrated the impact of using different attenuation models and the limited bandwidth on estimating the spectral source properties of the small and moderate-size events. This study does not attempt to propagate the model-related error quantitatively as Bindi et al. (2023) but to demonstrate how to reduce the uncertainties in the measurements by

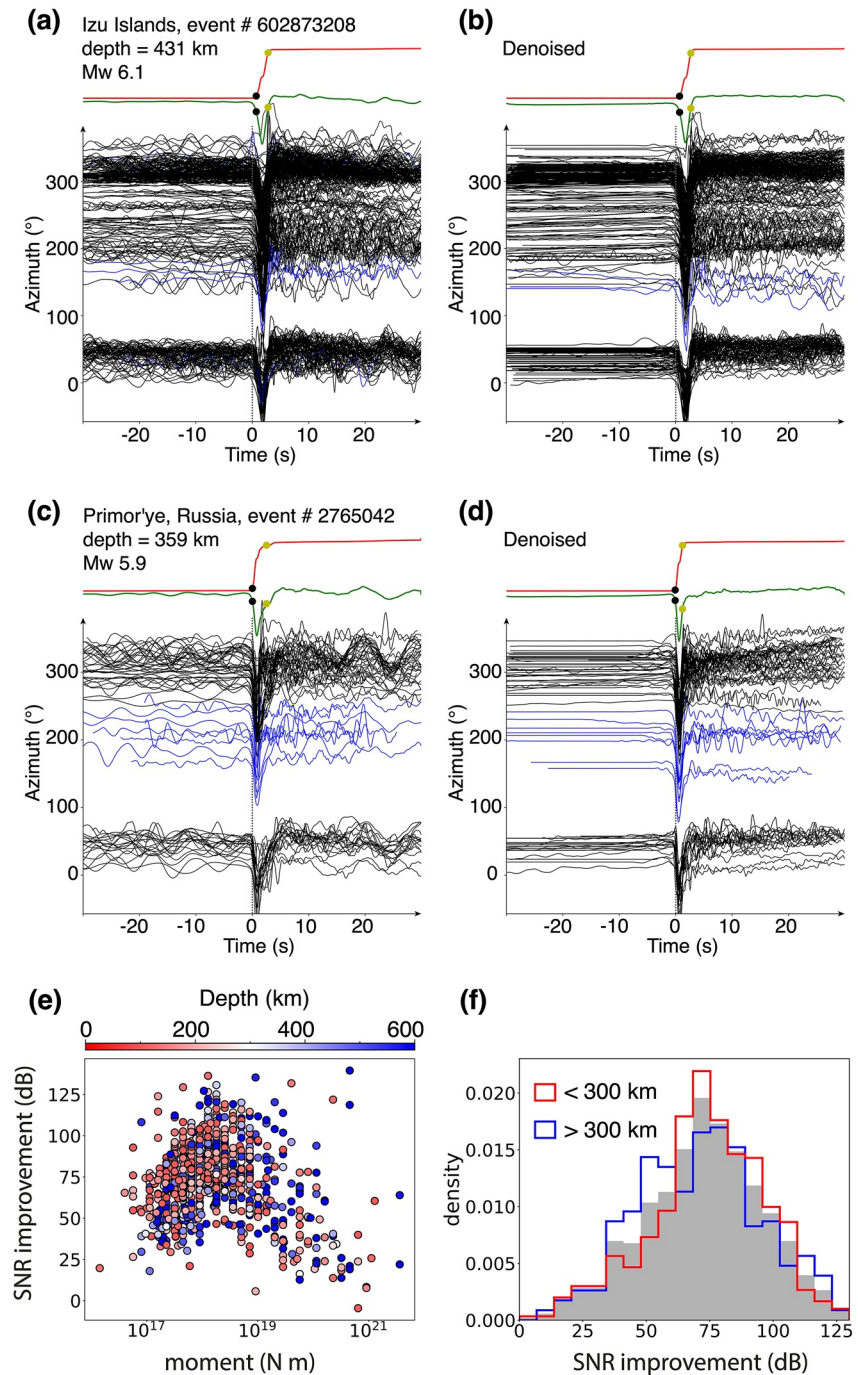


Figure 5. Denoising performance on real earthquakes. Denoising performance on two representative earthquakes deep earthquakes: the Mw6.1 2013 April 21 earthquake near the Izu Islands in Japan and the Mw5.9 2002 February 1 earthquake at Primor'ye in Russia. (a) and (c) show the original displacement waveforms, and (b) and (d) show the denoised waveforms. The waveforms are aligned with the peak amplitude, stretched based on the maximum cross-correlation coefficients, and sorted by azimuth relative to the epicenter. The blue waveforms are flipped in polarity for better visualization. The dashed line marks the onset of the P waves. The stacked displacement waveform is shown in green. The cumulative energy waveform shown in red is computed using the integral of the squared stacked velocity waveform. The black and yellow dots indicate the onset and termination time of the energy growth, which defines the duration. Denoising performance of overall deep earthquakes: (e) shows the increment of average SNR of each event by denoising as the function of the moment, color-coded by depth. (f) Shows the statistics of SNR increments for all deep earthquakes in this study (gray) and for intermediate-depth earthquakes (red) and deep-focus earthquakes (blue). The SNR improvement is defined by the increment $20 \log_{10}(\text{SNR})$.

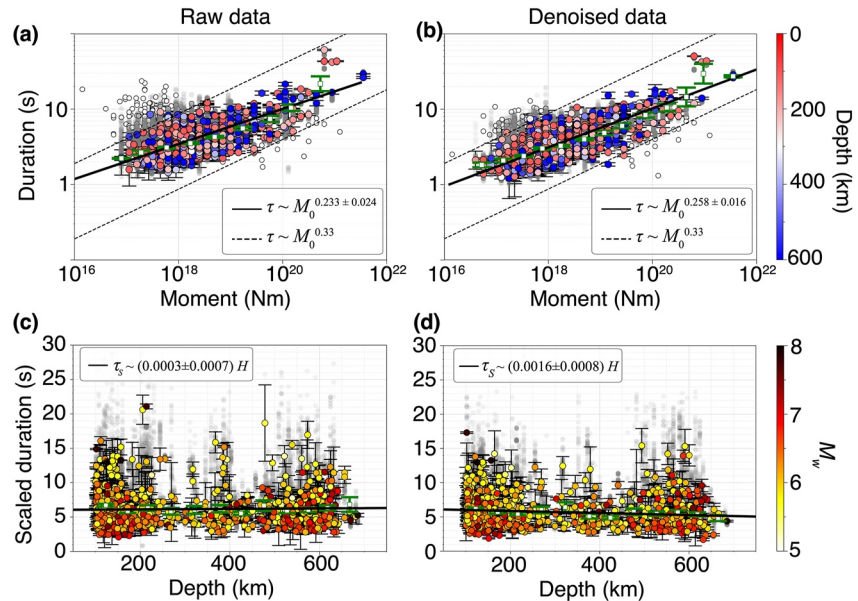


Figure 6. Durations scaling with magnitude and depth. The duration scaling coefficients derived from the raw data are shown in (a) and (c) in comparison to the duration scalings derived from the denoised data in (b) and (d). Moment-dependence: (a) and (b) the source duration rescaled by S-wave velocity is shown as bigger circles color-coded by depth and compared with the self-similar scaling relationships in dashed lines. The gray circles with 50% transparency are calculated by bootstrapping with stations, and the black error bars denote their standard deviations. The open black circles are the bootstrapped mean values of eliminated events due to the relatively unstable measurements (depth phases, stronger coda, insufficient azimuthal coverage, etc.). The black solid line indicates the optimal scaling relationship, estimated with the average duration (green squares) of 1,000 bootstrapped mean values over events (standard deviation as the green error bar) in each moment bin. The uncertainty of the final scaling is denoted in the legend. (c) and (d) The magnitude-scaled source duration (Equation 2) against depth and color-coded by the event magnitude. The green square, green error bar, gray circle, and black error bar are in the same fashion as (a) and (b). The best-fit scaling is shown as the black solid line with the slope uncertainty in the legend. H: Earthquake focal depth.

denoising low SNR data and how it does so by widening the frequency range of available data. Neely et al. (2020) found that large uncertainties exist for stress drop measurements and are not comparable between different studies of different tectonic settings. We analyze the variations of the global deep earthquake source properties based on the uniform model assumption and consistently compare them.

We explored a quantification of uncertainties in the source parameters by using bootstrapping over the stations for both the raw data and denoised data. While this quantifies “data consistency” over epistemic uncertainties, the variability in the data may represent the range of possible uncertainties due to SNR, frequency content, imperfect attenuation removal, and source directivity. The distribution of source parameters measured from the subsets, the corresponding standard deviations as well, and the deduced scaling relationships are shown together with the final measurements of the full data set in Figure 6.

In the following subsections, we select the denoised deep events with at least 20 data in at least six azimuthal bins (each of 45° width). This selection leads to **783** deep Mw5.0+ earthquakes for further analysis and ensures that the statistical properties of deep earthquakes are not biased by imperfect data coverage. This about doubles the number of events studied relative to Poli and Prieto (2016).

4.1. Source Duration

The event source duration is assumed to be the measured pulse width of the stacked P displacement waveform (we ignore the broadening of the pulse due to attenuation). This assumption is made because displacement seismograms are proportional to moment rate functions in the far field of an attenuation-free whole space. We first shift the time series using cross-correlation. We use the highest SNR trace as a reference and align all others using cross-correlation. We normalize the waveforms with their maximum amplitudes (flipping those with negative

polarity). We then stack the aligned and normalized traces for a first reference waveform. In a second iteration, we align the waveforms according to the first reference. We show these aligned and normalized waveforms in Figure 5.

In the second iteration, we take the stacked waveform as a reference to align each normalized trace again. We then stretch each normalized trace according to the reference using the stretching ratio that maximizes the Pearson coefficient between the stretched trace and the reference. We then stack the aligned and stretched pulses to obtain our improved stacked P-wave pulse.

We measure the source duration of the average from cumulative energy. We first take the derivative of the stacked displacement pulse (i.e., velocity), square it, and integrate it over time to compute the cumulative energy function. A typical cumulative energy function shows a flat-ramp-flat shape, where the time when cumulative energy rises corresponds to the source duration. We use the time when 5% and 90% of the total energy are reached to approximate the onset and termination of the event. The threshold choice was chosen to mitigate the artifact of the coda waves. All durations done in the time domain follow this calculation.

Because earthquake duration varies greatly with earthquake magnitude, we also calculate the scaled duration τ_s in a similar way to Houston et al. (1998) and Poli and Prieto (2014), using the following definition,

$$\tau_s = \frac{\beta}{\beta^{ref}} \left(\frac{M_0^{ref}}{M_0} \right)^{\frac{1}{3+\epsilon}} \tau, \quad (2)$$

where τ is the source duration, β is the shear-wave velocity at the event depth of the Preliminary Reference Earth Model (PREM) (Dziewonski & Anderson, 1981), and M_0 is the event seismic moment. M_0^{ref} is the reference moment 10^{19} N m and β^{ref} is the shear-wave velocity 4.4 km/s at the reference depth 170 km. Here, ϵ represents the departure from the self-similarity and is fit to the data (Houston et al., 1998; Kanamori, 2004; Poli & Prieto, 2014). The map view of the scaled duration is shown in Figure S2 of the Supporting Information S1.

We also measure duration as the inverse of the corner frequency. Section 4.4 discusses how we perform spectral fitting, extracting the corner frequency that is inversely proportional to the duration. We test this relation and show it in Supporting Information S1 (Figure S3).

The source duration of moderate-size earthquakes ($10^{16} < M_0 < 10^{19}$ N m) shows relatively higher variability than those of larger earthquakes ($M_0 > 10^{19}$ N m), possibly due to the limited number of large events or sensitivity to residual noise (Figures 6a and 6b). This increased variability at low magnitudes is typical of studies Allmann and Shearer (2009), Denolle and Shearer (2016), and Courboulès et al. (2016). As shown in Figure 6b, the source duration of the earthquakes of moments around 10^{18} N m (equivalent to $M_w 5.9$) ranges between 1 and 8 s, which is about an order of magnitude difference. The duration measurement taken as the inverse of the corner frequency exhibits similar variability (Figure S3 in Supporting Information S1).

Potential errors that introduce variability in the measurements could be attributed to depth phases of the shallowest deep earthquakes, which can be easily eliminated for short-duration events using a cut-off time window of 0–20 s following the first arrival but could be difficult to remove for long-duration events where the depth phases interfering with the direct phases.

We fit the observed $\log_{10}\tau \sim a \log_{10}M_0$ with linear regression, where the duration is corrected with the depth-dependent bulk properties (i.e., shear-wave velocity). We find that $a = 0.25 \pm 0.02$ matches best with the moderate- to large-magnitude earthquakes, and this represents the scaling $\tau \sim M_0^{0.25}$. The measurements of the inverse of corner frequency further confirm the scaling assuming $\tau = 1/f_c$ (see Figure S3 in Supporting Information S1). This scaling is similar to what has been found for intermediate-depth and deep-focus earthquakes (Allmann & Shearer, 2009; Poli & Prieto, 2016; Turner et al., 2022).

The depth dependence in scaled duration is well explained by the depth variations in material properties, or equivalently, that the scaled duration is depth independent. Given a reference magnitude of $M_w 6.6$, the scaled duration at a depth of 100–250 km has a mean value of about 5.5 s, while those at a depth of 500–600 km have a mean value of about 5.3 s. The mean scaled duration, when estimated from corner frequency (i.e., $1/f_c$), of the intermediate-depth and deep-focus events are both about 5.5 s. Similar variability of $1/f_c$ is found for the intermediate-depth and deep-focus events (2–12 s).

4.2. Directivity Effects

The rupture directivity alters the shape of far-field P-wave pulses by stretching or squeezing the seismic waveforms with ratios that vary with the azimuths and take-off angles away from the direction of rupture propagation. Directivity effects usually yield a shorter apparent duration and an enhanced high-frequency content in the direction of rupture propagation. These effects may be referred to as Doppler effects. When the earthquake rupture propagates in a unilateral direction, the Doppler effects are clear and asymmetric with respect to the direction of rupture. When the earthquake rupture propagates fast, as measured by the ratio of the rupture speed V_r to the velocity of the seismic wave propagation V_p , it enhances the contrast in apparent duration and magnifies Doppler effects.

Figure S4 in Supporting Information S1 illustrates the geometrical relation between the direction of rupture and the direction of the seismic ray taking off. We modify Equation 1 of Park and Ishii (2015) to express the apparent duration of the P-wave pulse at station i , τ_i :

$$\tau_i = \frac{L}{V_r} \left(1 - \frac{V_r}{V_p} \cos \theta_i \right), \quad (3)$$

where V_r is the average speed of a unilaterally propagating through rupture, L is the total length of rupture, V_p is the P-wave velocity at the source, and θ_i is the angle between the rupture propagation and ray take-off directions. Because V_r tends to be closer to the shear-wave speed V_s , directivity effects in P-wave pulses are typically less than observed in S-wave pulses. Based on the geometry between the rupture directivity and the seismic ray path (Figure S4 in Supporting Information S1), $\cos \theta_i$ is

$$\cos \theta_i = \sin \gamma_i \sin \beta + \cos \gamma_i \cos \beta \cos(\phi_i - \phi_r), \quad (4)$$

where the angle parameters are explained and illustrated in Figure S4 of the Supporting Information S1. Each source-station geometry provides a unique set of geometrical parameters. We know ϕ_i and γ_i from the earthquake and receiver location and τ_i from measurements. We need to find L , V_r , β , and ϕ_r . We perform a grid search for the four parameters. β is searched between $-\pi/2$ and $\pi/2$ with 36 grid points, ϕ_r is searched between 0 and 2π with 72 grid points, V_r is searched within $0 \sim V_p$ with 100 grid points and L is searched between $0.6 V_r \tau$ and $1.4 V_r \tau$ with 8 grid points.

To get apparent V_r and the direction of directivity, we need to measure τ_i . We measure the τ_i at each station using the stretching/squeezing ratio between the station-specific and the station-stacked displacement P waveforms. Then, we take the ratio between the relative pulse durations and the average source duration.

We select the events with at least 20 data in at least six azimuthal bins (each of 45° width). The ratio of the optimal rupture velocity of the events with the local S-wave velocity is referred to as the ‘‘Doppler ratio’’ because it is only relevant for unilateral moving ruptures. Here, we cannot determine the rupture velocity of a radially propagating rupture, but we can assess the circularity of the rupture propagation with the Doppler ratio. A high Doppler ratio indicates a rather unilateral rupture and a low Doppler ratio indicates a rather circular rupture. Our measured Doppler ratio (V_{rup}/V_s) is shown in Figure 7a. Most earthquakes in this analysis have an apparent unilateral rupture speed slower than 50% of the S-wave velocity. Hence, we draw our first conclusion that unilateral propagation is not the dominant mode of propagation of deep earthquakes. Rather, the crack model of radially propagating rupture might well suit our observations.

We report that the denoised waveforms yield a much-reduced variance among the station-specific Doppler ratio values. We attribute this to the enhanced cross-correlation coefficients of stretched P waves, contributing to a more precise estimation of the relative source durations.

Our result shows a significant correlation between the estimated V_{rup}/V_s and earthquake moment. The smaller earthquakes have a broad range of Doppler ratios between 0.0 and 0.8, with a mean value of 0.3 (Figure 7a). This means the equivalent unilateral rupture speeds of the moderate-size deep earthquakes are mostly lower than 30% of the S-wave velocity. The large deep earthquakes have a narrower range of Doppler ratio values between 0.0 and 0.4, with a mean value of 0.15. The decrease of the maximum Doppler ratio with the increasing moment may be related to (a) the weakening of material beyond the seismogenic width (i.e., the slab) or (b) the growing complexity of the rupture processes, which can be involved with multiple faults or multiple mechanisms during

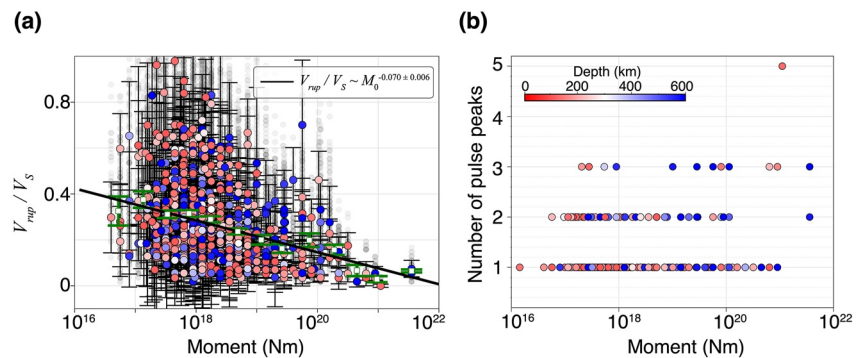


Figure 7. The Doppler effect of deep earthquakes analyzed in this study. (a) The equivalent unilateral rupture speed ratio to the S-wave velocity near the earthquake source is plotted to show the relation with the moment, color-coded by event depth. The green square and green error bars are the average value and the standard deviation of the bootstrapped mean Doppler ratios in each moment bin. The gray circle and the black error bar are the Doppler ratio using bootstrapped stations and the corresponding standard deviation for each event. The best-fit scaling relationship between the Doppler ratio and moment is shown as the black solid line with uncertainty denoted in the legend. (b) The number of peaks of the source time function in relation to seismic moment color-coded by event depth.

a single large deep event, leading to more homogeneous rupture propagation and a poorer representation of the directivity with the Doppler ratio.

We conduct statistical tests to demonstrate the significance of the difference between the distributions of the Doppler ratio at different depths. The null hypothesis is that the mean of the two distributions of Doppler ratios (depth ranges of 100–300 km and 300–700 km) are equal. We then obtain a *t*-score of 1.6 with an associated *p*-score of 0.11. Hence, this is not sufficient evidence to reject the null hypothesis. Therefore, Doppler ratios of earthquakes at the depth range of 100–300 km are statistically similar to that of earthquakes deeper than 300 km.

4.3. Earthquake Complexity With Subevents

Complex earthquake ruptures may comprise subevents that are bursts of moment release well separated in time (Antolik et al., 1999; Danré et al., 2019; Duputel et al., 2012; Houston et al., 1998; Ihmlé, 1998; Kikuchi & Fukao, 1987; Shi & Wei, 2020; Tibi et al., 2003; Tsai et al., 2005; Wei et al., 2013; Yin et al., 2021; Zhan, Kanamori, et al., 2014). We count the number of peaks of the stacked P-wave displacement for all deep earthquakes analyzed in this study. We use a peak detector function (`scipy.signal.find_peaks` in Python) and only search between the P-wave arrival time and the apparent duration. The data has been low-pass filtered below 4 Hz before integrating into displacements. We pick the subevent peaks from the stacked displacement over stations. We found that most events have between 1 and 3 subevents, as shown in Figure 7b. The waveform resolution (<4 Hz) is sufficient for *M*_w > 6 events and well below some *M*_w5.0–6.0 earthquakes. Three subevents are only detected for *M*_w > 5.5, and smaller events present fewer subevents (i.e., 1 or 2) as shown in Figure 7b. Larger earthquakes have a few and more subevents than smaller earthquakes, but overall, deep earthquakes are simpler ruptures with fewer subevents confirming Danré et al. (2019) and Yin et al. (2021) and the hypothesis that deep earthquakes are rather crack-like.

In the following section, we assume a simple source model to estimate the parameters in the frequency domain, which we justified using two metrics: rupture directivity and source pulse complexity quantified by the number of subevents. Our result shows that the large deep earthquakes are rather symmetric in terms of rupture propagation and contain fewer subevents compared to their shallow counterparts. This supported our choice of a simplified source model, even for large earthquakes.

4.4. Spectral Fitting

The far-field P wave displacement waveforms are an approximation to the moment-rate function. Their amplitudes are controlled by moment, radiation patterns, and geometrical spreading, mostly frequency-independent. The seismogram amplitudes are also affected by seismic attenuation, which considerably decreases the seismic

amplitudes at frequencies greater than 1 Hz. It is common in seismology to remove the attenuation effect by correcting the amplitudes in the frequency domain. We first transform the displacement time series to the Fourier amplitude spectrum using the package `mtspec` (Prieto, 2022; Prieto et al., 2009), which uses a multi-taper spectral analysis that is robust for short windows (Thomson, 1982). To correct for the attenuation of high-frequency energy for teleseismic P waves, we use the following equation,

$$\hat{S}(f) = \hat{U} e^{2\pi f t^*/2}, \quad (5)$$

where the attenuation parameter $t^* = 0.3$ s for the P waves that originate from the mantle (Poli & Prieto, 2016). f is frequency, $\hat{U}$ is the observed spectrum and $\hat{S}(f)$ is the corrected spectrum. To understand the trade-offs with attenuation, we add experiments using different attenuation parameters $t^* = 0.05, 0.15, 0.25, 0.35$, and 0.45 s, and summarize the effect in Figure S11 of the Supporting Information S1. We notice that when $t^* > 0.4$ s, the spectra (below 2 Hz) start to deviate from the theoretical source spectra, which is a sign of over-correction. Within the range of 0.05 – 0.35 s, the t^* does not change much the scaling between the spectral parameters and magnitude of Mw5.0+ deep earthquakes discussed later. This experiment validates that $t^* = 0.3$ s is overall a good value that leads to results consistent with the temporal measurements ($1/f_c \sim T$). We then scale each attenuation-corrected displacement spectra to one. To avoid biases of azimuthal distributions in the station coverage, we group the P-wave spectra into eight $\pi/4$ -wide azimuth bins. We first compute the average spectrum in each bin if there is data, then stack the spectra over azimuth bins, ignoring those without data. This procedure is to approximately correct the radiation pattern and geometrical spreading effects. We then level the stacked P spectra with the ISC catalog earthquake moment. Next, we use the following equation to model the source spectrum, assuming a Brune model (Brune, 1970).

$$\hat{S}'(f) = \frac{M_0}{1 + \left(\frac{f}{f_c}\right)^n}, \quad (6)$$

where the two parameters to find are the fall-off rate n and corner frequency f_c . The choice of a simple spectral shape is justified because of the low Doppler ratio and low complexity of the P-wave pulses. We perform fitting in the log-log space: log of amplitudes resampled on a log-frequency array. We then perform a grid search by minimizing the mean square residuals between the modeled and observed spectrum below 1 Hz. We limit the grid search to 2.5 Hz for the corner frequency, approximately the corner frequency (or inverse of duration) of an Mw5.0 earthquake based on the regional data analysis of intermediate-depth earthquakes by Prieto et al. (2013). A visual comparison between the optimal modeled spectra with the stacked spectra of the noisy and denoised P waves is shown in Figure S5 of the Supporting Information S1. The difference in spectral shapes between the synthetic and stacked spectra is reduced after denoising.

There are clear trade-offs when fitting for both fall-off rates n and corner frequencies. This problem was extensively discussed in the literature and widely recognized by the community (Denolle & Shearer, 2016; Eulenfeld et al., 2021; Huang et al., 2016; Kaneko & Shearer, 2015; Shearer et al., 2019; Trugman & Shearer, 2017). We address similarly this issue by solving for both parameter in the grid search and set a upper bound of 2.5 for the fall-off rate and 2.5 Hz for the corner frequency.

We now explore the effects of earthquake size on the shape of the observed and modeled spectra. We group the spectra in seven-magnitude bins by normalizing all spectra and leveling them to the bin central moment. We show the bootstrapped spectra in Figure 8. We average the logarithmic spectra amplitude in each magnitude bin by bootstrapping (selecting with replacement) 1,000 times the data. We obtained 1,000 averaged spectra, shown in Figure 8, and then averaged again for a single stacked spectrum per magnitude bin. We perform the same analysis for the original and the denoised seismograms.

The main results that can be interpreted are the variation of the corner frequencies with the seismic moment for the denoised seismograms (Figure 8b). We find a visual correlation that $M_0 \propto f_c^{-4}$, again supporting a deviation from a self-similar behavior. This result holds when considering the 783 individual estimates of f_c (Figure S3 in Supporting Information S1) and confirms the inverse relation between duration τ and moment, $M_0 \sim \tau^4$, illustrated in Figure 6b.

With the recognition that such noisy waveforms (Figure 8a) would be disregarded in seismological studies, we want to highlight the impact of including noise in the spectral fitting. Microseismic noise particularly biases the

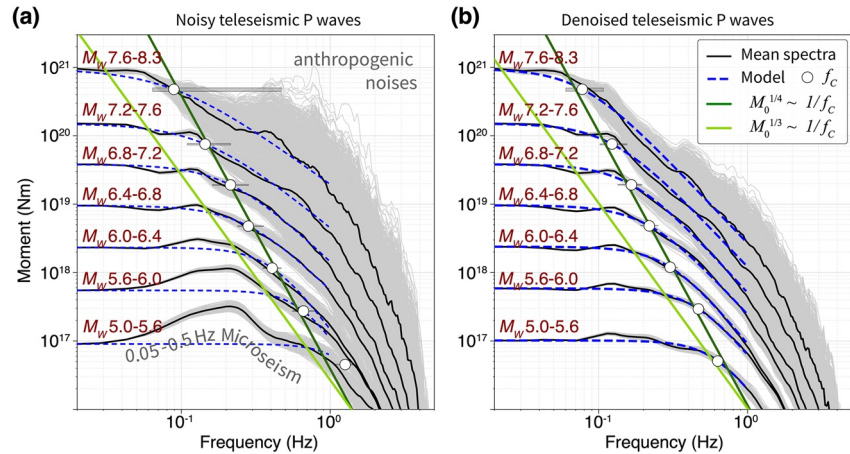


Figure 8. Spectra averaged in magnitude bins. (a) The noisy spectra are divided into seven magnitude groups, as indicated on the left, and bootstrapped in each group 1,000 times to compute the average spectra (gray). The median of the bootstrapped spectra mean (black lines) is well fit by the spectral model (blue dashed lines) after searching for the optimal corner frequency (white dots) and high-frequency fall-off rate. The gray bars indicate the uncertainties of the corner frequencies. The darker green line represents the scaling between corner frequency and moment fitted to the measurements, and the light green line represents the self-similarity model. (b) Same as (a) for the denoised waveforms.

retrieval of corner frequency for magnitude Mw5.0–6.5. Moreover, high-frequency noise biases the retrieval of the high-frequency fall-off rate (and thus corner frequency given the parameter trade-offs) of the larger earthquakes.

4.5. Stress Drop

As stated above, we justify using a circular, radially propagating crack model for deep earthquakes because (a) the source spectra fit well using a single-corner frequency model, (b) we infer the weak directivity effects, and (c) we find low complexity in the time-domain pulse. Crack models are modes of rupture where the fault slips behind the rupture front from the beginning of the fault slip until the earthquake fully arrests. We use the classic model of Brune (1970) later updated by Madariaga (1976) and Wang and Day (2017) to relate event duration and moment to stress drop $\Delta\sigma$:

$$\Delta\sigma = \frac{7}{16} M_0 \left(\frac{f_c}{0.35V_s} \right)^3, \quad (7)$$

where the geometrical parameter $7/16$ is used for a circular crack, the radius of the crack is estimated as $0.35V_s/f_c$. Here, the coefficient 0.35 is obtained in Wang and Day (2017) in dynamic rupture models of expanding cracks. We extract the shear-wave velocity V_s from the 1D PREM model (Dziewonski & Anderson, 1981). We show the values of stress drop in Figures 9c and 9d and Figures 10c and 10d, together with the uncertainty distributions obtained by bootstrapping the stations. We find a strong scaling of stress drops with earthquake magnitude but no variation with depth. We perform a linear regression $\log_{10}(\Delta\sigma) \sim a \log_{10} M_0$ using linear-least squares and find the exponent $a = 0.21 \pm 0.03$. The resulting strong scaling suggests that if the Mw5.0 earthquakes have a stress drop of about 1.8 MPa, the Mw7.5 earthquakes have a stress drop of 10 MPa. This scaling is slightly weaker than that found by Poli & Prieto (2016), though we generally find lower stress drops more consistent with global studies and crustal earthquakes (Allmann & Shearer, 2009), and using the time-domain duration estimate T would decrease the mean value of stress drop.

As expected from the non-typical scaling of duration with seismic moments, the scaling of stress drop with magnitude is strong (Figures 9c and 9d). We bootstrap the stress drop of the events in the moment bins, calculate average stress drops, perform a linear regression in the log-log space, and find a best slope of 0.21 ± 0.03 , such as $\Delta\sigma \sim M_0^{0.21 \pm 0.03}$. Furthermore, the scaling is stronger for earthquakes deeper than 300 km: “intermediate depth” earthquakes have a scaling $\Delta\sigma \sim M_0^{0.23}$ and “deep focused” earthquakes have a scaling of $\Delta\sigma \sim M_0^{0.26}$, as shown in Figure S6 of the Supporting Information S1.

Unsurprisingly, the variability in spectral shapes shown in Figure 8a yields a higher variability in corner frequency and, consequently, in estimated stress drop. The variability may be unreasonable and span four orders of magnitude higher than for the same waveforms but denoised using DenoTe. Therefore, our denoising technique has been essential

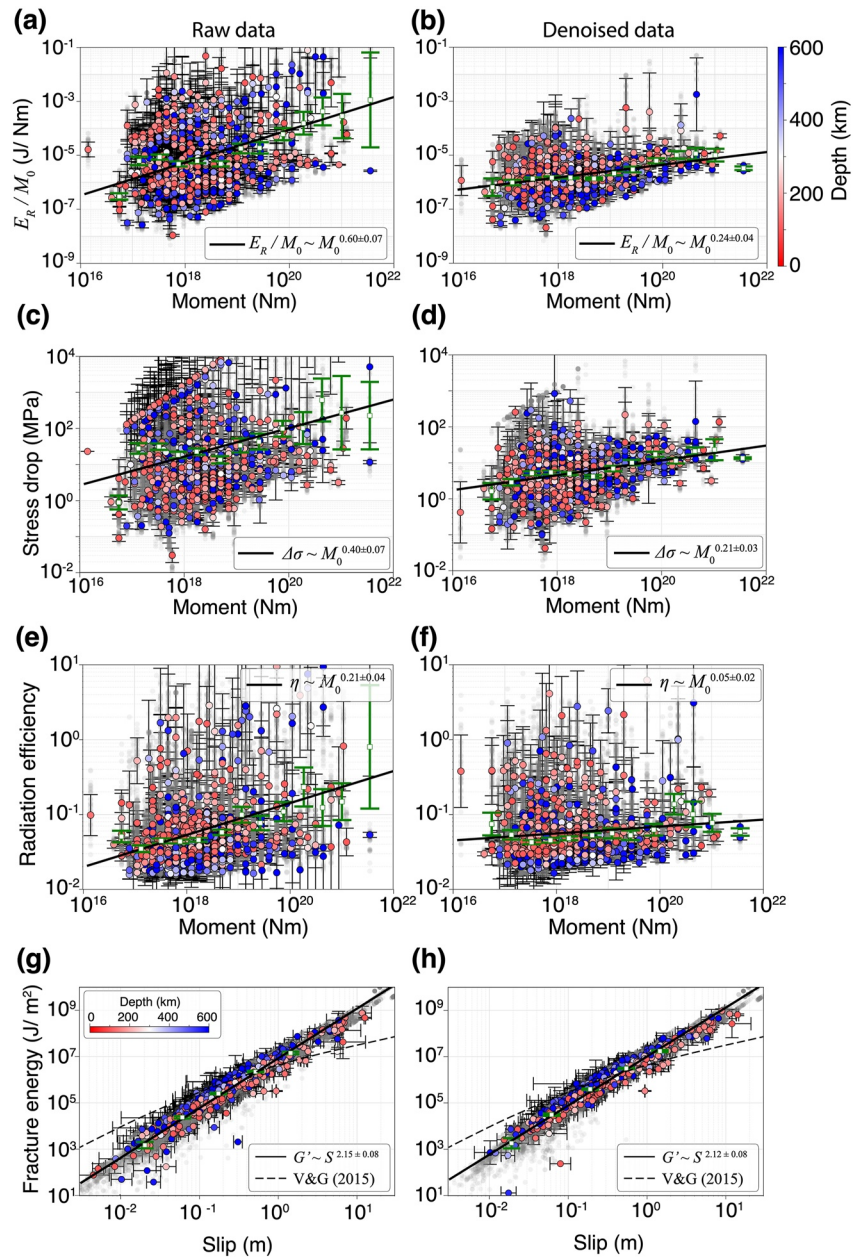


Figure 9. Radiated energy, Stress Drop, Radiation efficiency, Fracture energy, and earthquake size. (a), (c), (e), and (g) are derived from the raw data while (b), (d), (f), and (h) are derived from the denoised data. The source parameters in each subfigure are presented similarly: Bigger circles are individual events color-coded by depth. Each gray circle with 50% transparency and black error bar are measurements using bootstrapped stations and the corresponding standard deviation for each event. Each green square and error bar are the average value and the standard deviation of the bootstrapped mean over events in each moment bin. The optimal scaling relationship between the X and Y axes is the black solid line with slope and uncertainty denoted in the legend. (a) and (b) Moment-scaled radiated energy as a function of moment. (c) and (d) The scaling relationship between stress drop and moment. (e) and (f) Radiation efficiency as a function of moment. (g) and (h) The scaling relationship between earthquake average slip and fracture energy, in comparison with the dashed curves derived from the thermal-pressurization-based model by Viesca and Garagash (2015).

and provides more precise stress drop measurements and their scaling with magnitude. We calculate the stress drop using the duration estimates and find similar moment-dependence (Figure S7 in Supporting Information S1).

We do not see any strong dependence between stress drop and depth (Figures 10c and 10d). We measure an increased variability of the shallowest intermediate-depth earthquakes, which may indicate that we have less

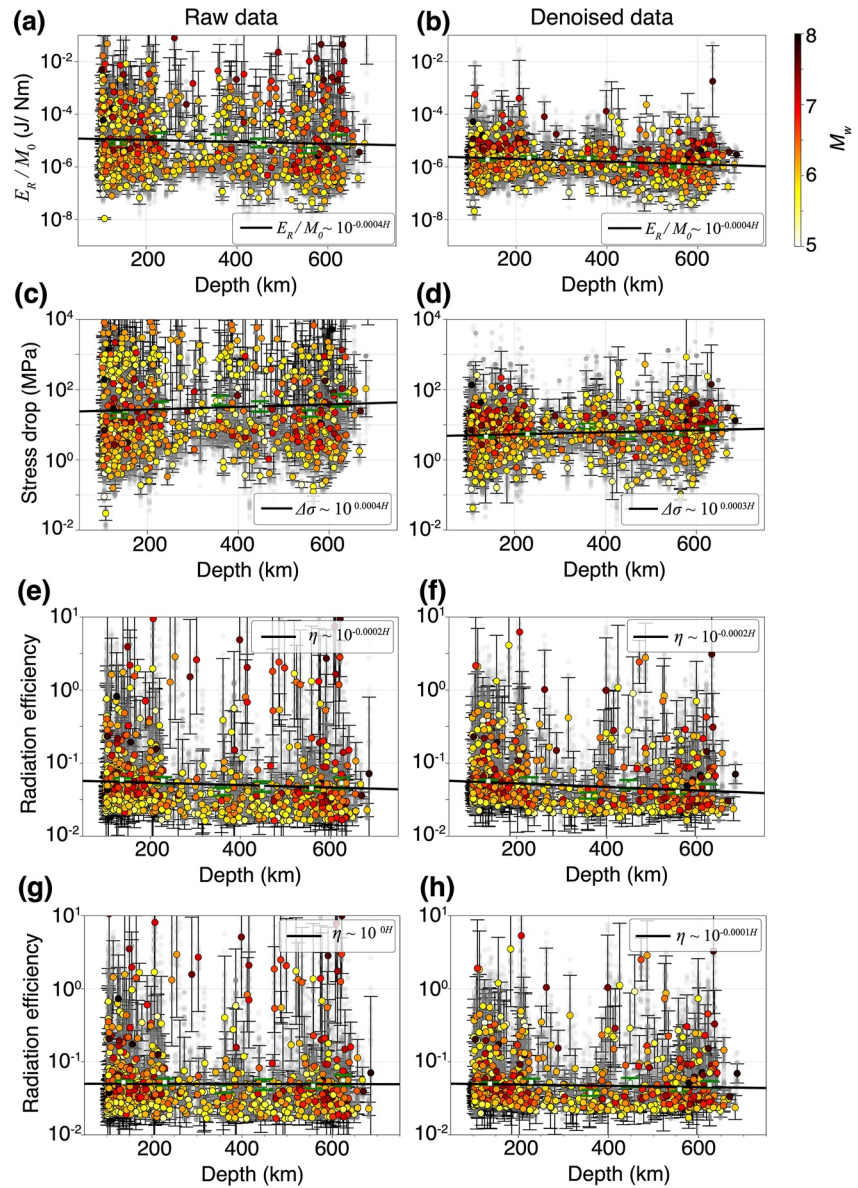


Figure 10. Radiated energy, Stress Drop, Radiation efficiency and earthquake depth. (a), (c), (e), and (g) are derived from the raw data, while (b), (d), (f), and (h) are derived from the denoised data. The source parameters in each subfigure are presented similarly: Bigger circles are individual events color-coded by magnitude. Each gray circle with 50% transparency and black error bar are measurements using bootstrapped stations and the corresponding standard deviation for each event. Each white square and green error bar are the average value, and the standard deviation of the bootstrapped mean over events in each depth group. The optimal scaling relationship between the X and Y axes is shown as the black solid line. Using the PREM velocity model: (a) and (b) Scaled radiated energy as a function of depth, (c) and (d) stress drop and depth, (e) and (f) radiation efficiency and depth. Using AK135 velocity model (g) and (h) radiation efficiency and depth.

stable duration measurements for the shallowest earthquakes (some depth phases may leak in our measurements), a greater sensitivity of the measurements to unknown attenuation effects, or may indicate a greater heterogeneity in source properties of shallow earthquakes. Overall, a depth-varying attenuation correction does not seem as necessary as that of the observed stress drop of crustal earthquakes (Abercrombie et al., 2021). Vallée (2013) found the constant strain drop with depth better fits the data. We scale the source duration using Equation 2, a different approach from Vallée (2013), based on the assumption of constant stress drop with depth. However, since the density and S-wave velocity vary by some moderate amount, we can not discriminate between constant stress drop and constant strain drop.

4.6. Radiated Energy

Next, we estimate the radiated energy of these earthquakes using the denoised waveforms. The kinetic energy of the radiated P wave can be estimated by integrating the squared P-wave velocity spectrum. We were partially motivated to measure if ML-denoising affected the waveforms over a range of frequencies to which radiated energy is particularly sensitive. We estimate the radiated P-wave energy using,

$$E_P = \frac{2\pi M_0^2 \langle R_p^2 \rangle}{\rho V_p^5} \int_0^\infty [f \hat{S}(f)]^2 df, \quad (8)$$

where, $\langle R_p^2 \rangle = 4\pi/15$ is the squared P-wave radiation pattern coefficient averaged over the double-couple focal sphere assuming the uniform shape of source spectra $\hat{S}(f)$, V_p is the P wave velocity at the location of the source and ρ is the density of the material at source. The seismic moment M_0 is calculated from moment-magnitude.

With the radiated energy, we can further calculate the apparent stress (see Figure S8 in Supporting Information S1) by

$$\sigma_a = \mu E_R / M_0, \quad (9)$$

The shear modulus μ is calculated with the shear-wave velocity of the PREM model.

In general, the observed spectra well match the model $\hat{S}'(f)$ in Equation 6 below 1.5 Hz (see Figure 8; Figure S11 in Supporting Information S1). Higher than 1.5 Hz, the observed spectra have a steeper fall-off than the model, which implies that attenuation may be frequency-dependent. Ide and Beroza (2001) has indicated that the source spectrum at frequencies higher than 10 times the corner frequency only accounts for less than 10% of the total energy. Hence, we separate the integration in Equation 8 in two parts: observed spectra integrated below 1 Hz and modeled spectra integrated over 1–4 Hz.

Similar to (Boatwright & Choy, 1986; Convers & Newman, 2011; Denolle & Shearer, 2016; Poli & Prieto, 2016), we scale the S energy using the ratio $E_S = 3V_p^5/2V_s^5 E_P$. Several assumptions are required to apply this ratio. First, S waves are assumed to have the same spectral shape as P waves. Second, we assume that the focal mechanism of the source is strictly a double couple, which is questionable for deep earthquakes (Frohlich, 1989; Green & Houston, 1995; Knopoff & Randall, 1970), and that we are sampling the whole focal sphere. Third, we assume the ratio between P and S waves found in the PREM velocity model.

We find that radiated energy also scales strongly with the seismic moment, with an exponent of 1.24. Figures 9a and 9b show the moment scaled energy E_R/M_0 as a function of moment. By performing linear regression $\log_{10}(E_R/M_0) \sim a \log_{10} M_0$, we obtain the exponent $a = 0.24 \pm 0.04$. Such scaling is expected from the scaling of corner frequency with earthquake magnitude because of the abnormally higher corner frequency of larger earthquakes, within which seismic energy concentrates. Typical self-similar concepts of earthquake scaling promote the idea that scaled energy, E_R/M_0 is constant (Baltay et al., 2010; Convers & Newman, 2011; Venkataraman & Kanamori, 2004), though Denolle and Shearer (2016) found the scaled energy tended to be constant regardless of the fault geometry.

We show the scaling relationships between radiated energy and depth derived from the noisy and denoised P waves in Figures 10a and 10b, respectively. Similar to the moment dependence, denoising reduces the variability of the radiated energy measurements but does not alter the general trend of the scaling.

4.7. Radiation Efficiency

Considering the simplified slip-weakening model of fault strength, we also calculate the apparent radiation efficiency introduced by Venkataraman and Kanamori (2004), also well explained and discussed in Abercrombie and Rice (2005), Noda and Lapusta (2013), and Lambert et al. (2021). We use the definition of radiation efficiency:

$$\eta_R = \frac{2\mu E_R}{\Delta\sigma M_0}, \quad (10)$$

where the shear modulus μ is calculated with the shear-wave velocity of the PREM model, seismic moment M_0 is calculated from moment-magnitude, radiated energy E_R and stress drop $\Delta\sigma$ are measured above.

We find low radiation efficiency at about 0.05, similar to other studies on deep earthquakes (Poli & Prieto, 2016; Prieto et al., 2013; Wiens, 2001). These values are typically much lower than those reported for crustal earthquakes (Lambert et al., 2021; Prieto et al., 2017; Singh et al., 2004; Venkataraman & Kanamori, 2004; Zollo et al., 2014). Noda and Lapusta (2013) and Lambert et al. (2021) suggested that radiation efficiency inferred from seismic observations tends to be overestimated as the seismological stress drop estimate is likely to be underestimated (Noda & Lapusta, 2013). Together with these potential biases, our results suggest deep earthquakes have much lower radiation efficiency than crustal ones.

We observe a weak moment-dependence of radiation efficiency (Figure 9f), with a slope of 0.05 ± 0.02 by a linear regression of the results obtained from the denoised data. This is also implied by the slight difference in scaling found for radiated energy and stress drop. Visually, there is greater variability of radiation efficiency for smaller magnitude earthquakes, which can be attributed to greater variability in corner frequency.

To further study the relationship between the radiation efficiency and source depth, we calculate the average radiation efficiency within each small depth interval (see Figures 10e and 10f). The shallowest earthquakes (100–250 km) have average radiation efficiencies about 30% higher than those of the events at greater depth. We can rule out attenuation effects: we have assumed a unique attenuation correction. Thus, it is possible that we over-corrected the deeper earthquake signals relative to shallower earthquake signals, which would give an apparently higher radiated energy. Because radiation efficiency as calculated in Equation 10 is effectively proportional to $1/V_p^3$, uncertainties from this ratio due to our choice of velocity depth profile can explain a portion of the depth-dependence. Nevertheless, our conclusions remain unchanged when using the AK135-f velocity model (Kennett et al., 1995; Montagner & Kennett, 1996). See Figures 10f and 10h; Figure S10 in Supporting Information S1 for comparison.

4.8. Fracture Energy

Fracture energy is the energy spent to create a fracture surface. We use the definition of the energy budget in Kanamori and Rivera (2006) for slip-weakening models of earthquakes to estimate the fracture energy from our seismic observables, stress drop, and scaled energy:

$$G' = \frac{1}{2}(\Delta\sigma - 2\sigma_a)S, \quad (11)$$

where σ_a is referred to as apparent stress and S is the average slip of the ruptured area that is calculated in an elliptical or circular model as $S = M_0 / [\mu\pi(0.35V_S\tau)^2]$. In this example, we use τ as our time-domain duration estimate. The spectral parameter 0.35 is obtained from Wang and Day (2017) assuming an expanding crack. It should be noted that the fracture energy can be underestimated in the case of undershoot, where the fault is weakened to a low friction level dynamically and recovers to higher friction when the slip stops (Viesca & Garagash, 2015). We show the estimated values in Figures 9g and 9h.

In general, visual inspection Figures 9h–9i suggests that deeper earthquakes exhibit slightly higher fracture energy, discussed earlier, with a slightly lower radiation efficiency relative to intermediate-depth earthquakes. Overall, both intermediate-depth and deep earthquakes share a similar relation between fracture energy and slip. This further suggests that their energy budget are similar despite the possible and diverse mechanisms discussed in Zhan (2020).

Typical scaling between observed fracture energy and average slip is $G' \sim S^2$, which is overall satisfied with our observations. By performing linear regression $\log_{10}(G') \sim a \log_{10} S$, we obtain the exponent $a = 2.12 \pm 0.08$ using the results obtained from the denoised data (Figure 9h). A similar scaling relationship is derived from the raw data (Figure 9g). This is consistent with the inference from Abercrombie and Rice (2005). For shallower earthquakes, Viesca and Garagash (2015) found a change in scaling for larger earthquakes that could be modeled using dynamic weakening mechanisms such as flash heating (Rice, 2006) and thermo-pressurization of fluids (Marguin & Simpson, 2023; Noda & Lapusta, 2013). In contrast to the inferred behavior of shallower earthquakes (Viesca & Garagash, 2015), our results suggest no strong dynamic weakening mechanisms.

The overall low radiation efficiency of moderate- to large-size deep earthquakes implies that the fault weakening is likely to be persistent during the slip growth so that fracture energy remains at a high level.

5. Discussion on the Properties of Deep Earthquakes

The weak directivity is a distinct feature of deep earthquakes, implying the relatively homogeneous stress states in the mantle or more diffusive rupture mechanisms. On average, we find Doppler ratios of 0.1–0.4 for $M_w > 7$ deep earthquakes, corresponding to 0.5–2.2 km/s apparent unilateral rupture speed, assuming an average S-wave velocity of 4.5–5.5 km/s. This is consistent with the slow rupture speed observed for large, deep earthquakes. Beck et al. (1995) derived a slow rupture speed (1–2 km/s, 636 km) for the 1994 $M_w 8.3$ Bolivian earthquake. Park and Ishii (2015) derived the average rupture speed for the 2012 $M_w 7.7$ (2.7 km/s, 583 km) and 2013 $M_w 8.3$ (1.4 km/s, 602 km) earthquakes in the Sea of Okhotsk region. Warren and Shearer (2006) studied the global deep moderate-to-large earthquakes during 1988–2000 and found slow rupture speed in most earthquakes. Prieto et al. (2017) obtained a best-fit slow unilateral and sub-horizontal rupture directivity (1.3 km/s) of the 2013 $M_w 4.8$ Wyoming earthquake (75 km). Díaz-Mojica et al. (2014) used an elliptical patch approach to study the 2011 $M_w 6.5$ Guerrero, Mexico earthquake (62 km) and found a slow rupture (0.5 km/s). Mirwald et al. (2019) also found a slow rupture (0.34 km/s) during the 2017 $M_w 7.1$ earthquake (57 km) in the Cocos plate beneath central Mexico. In contrast, Zhan, Helmberger, et al. (2014) used the duration after EGF correction and obtained a rupture speed above the local V_s for the $M_w 6.7$ Sea of Okhotsk earthquake (642 km), implying a very different rupture process relative to the nearby 2013 $M_w 8.3$ Okhotsk Earthquake. This may be confirmed by the larger variability of Doppler ratios we find for $M_w 5.0$ –6.9 earthquakes.

The moderate-magnitude earthquakes ($10^{16} < M_0 < 10^{19}$ Nm) have source dimensions comparable to the width of the subduction zone slab core. Within the core, frictional conditions may be more favorable for dynamic rupture, given the potentially elevated pore pressure due to mineral phase transformation (dehydration or compaction) or pre-existing slab faults. The larger-magnitude earthquakes have a greater spatial extent and, therefore, can further propagate into the surrounding mantle, which could have a less heterogeneous structure than the slab and considerably less water content. The distinct environments where these earthquakes reside may lead to scale-dependent Doppler ratios. The colder slab core may provide favorable conditions for small but faster rupture growth, while the surrounding warm material may be involved with a more dissipative and slower rupture.

Deep earthquakes have shorter source duration and thus higher corner frequencies than shallow earthquakes due to increased rigidity with depth (Vallée, 2013). The magnitude-duration scaling $M_0 \sim \tau^4$ that we measured from the denoised P waves is consistent with previous studies (Poli & Prieto, 2014). The corner frequency of deep earthquake displacement seismograms of direct P waves obtained from fitting Brune's models follows the same scaling with seismic moment ($M_0 \sim f_c^{-4}$), which are consistent with the time-domain measurements. The difference between this scaling and that found for shallow earthquakes (Allmann & Shearer, 2009) suggests that the rupture area and slip scaling are not self-similar.

Given the moment-duration scaling, we infer that stress drop increases with seismic moment. Early studies on the topic reported weak stress drop scaling (Frohlich, 2006), while some recent studies based on a larger number of stations and wider frequency band have found evident scaling (Poli & Prieto, 2016; Prieto et al., 2013). We obtain a similar moment-scaling of stress drop $\Delta\sigma \sim M_0^{0.21}$ for $M_w 5.0+$ earthquakes at a 100–700 km depth range. The signature of self-similar behaviors is rather complex for shallow earthquakes and is subject to great debate. It is either claimed by observations/inference of source duration and moment (Allmann & Shearer, 2009; Courboux et al., 2016), or re-interpreted as non-circular faults Denolle and Shearer (2016), or not found at all (Bindi et al., 2023; Trugman & Shearer, 2017). Cocco et al. (2016) compared stress drop estimates from different tectonic settings and using different methodologies to confirm the large variability up to three orders of magnitude (0.1–100 MPa, similar to the range in Figures 9c and 9d) for a broad range of seismic moment ($-8 < M_w < 9$), and reported no evident scaling of stress drop with earthquake size. While we cannot speculate on the scaling of stress drop with magnitude outside of our observational range, previous studies have found weak to no scaling of stress drop with smaller earthquakes in regional studies (Tian et al., 2022). What mechanism would change the scaling, whether it is geometrical or rheological, remains to be explored.

The radiation efficiency of deep earthquakes mainly ranges between 1% and 10%, much lower than that of shallow large events (25% by Kanamori and Brodsky (2004)). The low radiation efficiency and high-stress drop of these deep earthquakes could also be explained by substantial shear heating, similar to the interpretation of Prieto et al. (2013). We have ignored 3D velocity and attenuation models, which significantly impact the high-frequency content of the P-wave displacement, which should be incorporated in future work. We note that recent studies

have found much greater radiation efficiency (Turner et al., 2022). Therefore, our global study does not always represent the entire breadth of earthquake behavior.

Despite the argument that different mechanisms may enable intermediate-depth earthquakes and deep-focus (Zhan, 2020), they show similar characteristics in terms of magnitude scaling with duration, static stress drop, and radiated energy. The lack of depth variations in these parameters may also indicate that similar mechanisms govern the earthquakes in the two depth ranges. We note that the stress drop-magnitude scaling (power law of exponent 0.21) and the low median radiation efficiency (0.05) of both intermediate-depth and deep-focus earthquakes are similar to the result of Prieto et al. (2013). This indicates that the source processes of deep earthquakes could be dissipative and translate a small portion of static stress drop into high-frequency radiation. Hence, this study further extends the possibility of thermal runaway mechanism from the intermediate-depth earthquakes to the deep-focus events.

The study based on data from shallow earthquakes (Abercrombie & Rice, 2005) suggests the frictional strength decreases more rapidly in the initial stage of rapid slip and then decreases more slowly at larger cumulated slip ($\sigma_f(S) \propto -S^{0.28}$). Deep earthquakes show a more uniform decay rate of friction over slip distance ($\sigma_f(S) \propto -S^1$). Based on the scaling of fracture energy and average slip, deep earthquakes may not favor the dynamic weakening mechanism of thermal pressurization mechanism, Viesca and Garagash (2015) proposed to dominate for shallow events (Figures 9g and 9h). Alternative mechanisms may include flash heating and even melting, which require persistently high fracture energy for larger earthquakes. On the other hand, thermal pressurization may be greatly limited for deep earthquakes because of the depleted water or fluid at the depth range, especially if the earthquakes propagate in the mantle. Nonetheless, other mechanisms, such as shear heating, may be invoked to explain the large fracture energy and slow rupture propagation.

It appears difficult to invoke single mechanisms proposed for deep earthquakes (phase transformation, dehydration embrittlement, shear heating) to explain whole event dynamics. Our measurements of source dynamics favor the interpretation of dissipative shear heating as a dominant mechanism at the source, though dissipative mechanisms do not favor nucleation. Instead, the dual mechanism proposed by Zhan (2020) is practical and may explain the combination of dynamic nucleation and dissipative propagation. Besides, two nucleation mechanisms can be invoked to differentiate between intermediate-depth and deep-focused earthquakes. The intermediate-depth earthquakes may be initiated by dehydration embrittlement, and the deep-focus earthquake may be triggered by transformational faulting. As the rupture grows in size, thermal runaway takes over, leading to a large portion of stress drop being dissipated near the source. Due to the diffusive nature of heat transmission, shear heating allows for dynamic rupture, even if it's inefficient at radiating waves.

Deep earthquakes generally have relatively simple rupture processes compared to crustal earthquakes because of the fewer subevents identified from their source time functions. This feature may favor the fact that deep earthquakes tend to start on the faults with preferred orientation (e.g., along the metastable olivine wedge or along the pre-existing intra-plate faults) and develop with smooth propagation. This starting phase may be related to a relatively faster unilateral rupture speed (Zhan, Helmberger, et al., 2014). As the rupture is growing to a certain extent, the smooth propagation with the preferred fault orientation could be replaced with a slower and dissipative phase, which probably has a complex fault orientation (e.g., the 1994 Bolivia earthquake interpreted by Zhan, Kanamori, et al. (2014)).

Our neural networks can be easily generalized to other seismic waves with different window lengths and sampling rates. The fully connected layer between the shallow and deep kernels is adjustable, with higher learning capability for larger input sizes. Hence, the same architecture can be effectively applied to other seismic phases with minor modifications. Therefore, the general framework we developed in this study is of great potential to be applied to different types of research. An extension of this work could be extending the analysis for shallow earthquakes, which are still offshore and have coverage on island stations that are polluted with microseismic noise. The denoised waveform can provide Green's functions with better azimuthal coverages.

Another widely employed research is receiver function studies that rely on the data quality of the three-component teleseismic seismograms. With the P wave denoiser, the secondary phases can better stand out from the strong noise, so it provides many-fold more data recordings: 135,265 traces of Mw5.0–5.5 deep earthquakes were selected based on SNR > 8 after denoising, while only 3,118 of them could have been used with the same SNR criterion without denoising. We show the overall improvement for individual deep earthquakes in Figures 5e

and 5f. Furthermore, the application of our “DenoTe” to regional seismic networks would greatly benefit the real-time phase picking for larger-scale earthquake monitoring and enhance the accuracy of both the travel-time-based and waveform-based tomography studies.

6. Conclusion

This study demonstrates that machine learning can be included as data pre-processing to enhance our observation capabilities for earthquake source characterization. The demonstration uses deep earthquakes as an example because they already have relatively “clean” seismograms. Our ML denoising considerably improved the volume of data with a sufficiently good signal-to-noise ratio and an accurate wiggle-to-wiggle reconstruction over a broad range of frequencies, especially for smaller earthquakes. We doubled the number of events studied and considerably added independent observations (e.g., station waveforms) to each earthquake. We have demonstrated that broadband signals can be recovered using time-domain ML processing.

Our analysis of deep earthquakes is an update from the Poli and Prieto (2016) analysis, whereby we include more events of smaller magnitudes and expand beyond the analysis of scaling, depth dependence, energy budget, and earthquake complexity. We confirm the results of other studies (Poli & Prieto, 2016) that have found a strong scaling of stress drop and scaled energy with earthquake magnitude, which suggests weakening mechanisms stronger with earthquake size. It should be noted that this work has so far only validated the scaling relationships for deep earthquakes with $M_w > 5.0$. This result does not contrast much with the recent body of research published on shallow earthquakes (see references therein).

The lack of directivity effects and low complexity found for intermediate and deep earthquakes suggests that these events are rather crack-like and confined ruptures. We generally find that typical stress drops of 1–10 MPa and low scaled energy (10^{-5} J/Nm) have relatively low directivity, yielding low radiation efficiency and high fracture energy. While dynamic mechanisms may be at play for larger earthquakes, the rupture propagation of intermediate and deep earthquakes is dissipative.

There remain limitations to this work. Our preliminary test on S wave data was inconclusive because generating the data set of “clean” S waves is tedious and because S waves are much more depleted in high frequency than can be corrected for by a frequency-constant t^* model. There are clearly opportunities to incorporate ML denoising in other earthquake studies, such as receiver functions and finite source inversions.

Data Availability Statement

The software package for denoising is developed using PyTorch. It is named “DenoTe” and can be accessed from <https://github.com/qibinshi/TeleseismicDenoiser> and <https://zenodo.org/record/7807794> (Shi, 2023). We use data from the 1078 networks of the FDSN archive. The digital object identifier (DOI) of all 1078 networks can be found in Table S1. The minimally pre-processed seismic data used for training the neural network can be accessed at https://dasway.ess.washington.edu/shared/qibins/Psnr25_lp4_2000-2021.hdf5 (Shi & Denolle, 2023), and the waveform data and metadata for the deep earthquake analysis can be accessed at https://dasway.ess.washington.edu/shared/qibins/deepquake_M5_6_data_metadata.zip (Shi & Denolle, 2023). The earthquake catalog for selecting the waveform data is downloaded from ISC <http://www.isc.ac.uk/>. The DOIs of the seismic network involved in this study are saved as a ZIP file. The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to waveforms and metadata (last accessed July 2022).

Acknowledgments

We thank Editor Rachel Abercrombie, the Associate Editor, reviewer Daniel Trugman, and an anonymous reviewer for their constructive remarks that improved our manuscript. We acknowledge the National Science Foundation (CAREER award EAR 2124722) for supporting this research. The authors thank Jiuxun Yin for the discussions about his denoising.

References

- Abercrombie, R. E. (2021). Resolution and uncertainties in estimates of earthquake stress drop and energy release. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 379(2196), 20200131. <https://doi.org/10.1098/rsta.2020.0131>
- Abercrombie, R. E., & Rice, J. R. (2005). Can observations of earthquake scaling constrain slip weakening? *Geophysical Journal International*, 162(2), 406–424. <https://doi.org/10.1111/j.1365-246X.2005.02579.x>
- Abercrombie, R. E., Trugman, D. T., Shearer, P. M., Chen, X., Zhang, J., Pennington, C. N., et al. (2021). Does earthquake stress drop increase with depth in the crust? *Journal of Geophysical Research: Solid Earth*, 126(10), e2021JB022314. <https://doi.org/10.1029/2021JB022314>
- Abers, G. A., Nakajima, J., van Keken, P. E., Kita, S., & Hacker, B. R. (2013). Thermal–petrological controls on the location of earthquakes within subducting plates. *Earth and Planetary Science Letters*, 369–370, 178–187. <https://doi.org/10.1016/j.epsl.2013.03.022>
- Allmann, B. P., & Shearer, P. M. (2009). Global variations of stress drop for moderate to large earthquakes. *Journal of Geophysical Research*, 114(B1), B01310. <https://doi.org/10.1029/2008JB005821>

- Antolik, M., Dreger, D., & Romanowicz, B. (1999). Rupture processes of large deep-focus earthquakes from inversion of moment rate functions. *Journal of Geophysical Research*, 104(B1), 863–894. <https://doi.org/10.1029/1998JB900042>
- Baltay, A., Prieto, G., & Beroza, G. C. (2010). Radiated seismic energy from coda measurements and no scaling in apparent stress with seismic moment. *Journal of Geophysical Research*, 115(B8), B08314. <https://doi.org/10.1029/2009JB006736>
- Beck, S. L., Silver, P., Wallace, T. C., & James, D. (1995). Directivity analysis of the deep Bolivian earthquake of June 9, 1994. *Geophysical Research Letters*, 22(16), 2257–2260. <https://doi.org/10.1029/95GL01089>
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, 81(3), 530–533. <https://doi.org/10.1785/gssrl.81.3.530>
- Bindi, D., Spallarossa, D., Picozzi, M., Oth, A., Morasca, P., & Mayeda, K. (2023). The community stress-drop validation study—Part I: Source, propagation, and site decomposition of Fourier spectra. *Seismological Research Letters*, 94(4), 1980–1991. <https://doi.org/10.1785/0220230019>
- Boatwright, J., & Choy, G. L. (1986). Teleseismic estimates of the energy radiated by shallow earthquakes. *Journal of Geophysical Research*, 91(B2), 2095–2112. <https://doi.org/10.1029/JB091iB02p02095>
- Boneh, Y., Schottenfels, E., Kwong, K., van Zelst, I., Tong, X., Eimer, M., et al. (2019). Intermediate-depth earthquakes controlled by incoming plate hydration along bending-related faults. *Geophysical Research Letters*, 46(7), 3688–3697. <https://doi.org/10.1029/2018GL081585>
- Brudzinski, M. R., Thurber, C. H., Hacker, B. R., & Engdahl, E. R. (2007). Global prevalence of double Benioff zones. *Science*, 316(5830), 1472–1474. <https://doi.org/10.1126/science.1139204>
- Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research*, 75(26), 4997–5009. <https://doi.org/10.1029/JB075i026p04997>
- Chang, S., Yu, B., & Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. In *IEEE Transactions on Image Processing* (Vol. 9(9), pp. 1532–1546). <https://doi.org/10.1109/83.862633>
- Chen, Y., Savvaadis, A., Fomel, S., Chen, Y., Saad, O. M., Wang, H., et al. (2022). Denoising of distributed acoustic sensing seismic data using an integrated framework. *Seismological Research Letters*, 94(1), 457–472. <https://doi.org/10.1785/0220220117>
- Chu, S., Beroza, G. C., & Ellsworth, W. L. (2019). Source parameter variability of intermediate-depth earthquakes in Japanese subduction zones. *Journal of Geophysical Research: Solid Earth*, 124(8), 8704–8725. <https://doi.org/10.1029/2019JB017592>
- Cocco, M., Tinti, E., & Cirella, A. (2016). On the scale dependence of earthquake stress drop. *Journal of Seismology*, 20(4), 1151–1170. <https://doi.org/10.1007/s10950-016-9594-4>
- Convers, J. A., & Newman, A. V. (2011). Global evaluation of large earthquake energy from 1997 through mid-2010. *Journal of Geophysical Research*, 116(B8), B08304. <https://doi.org/10.1029/2010JB007928>
- Courbouloux, F., Vallée, M., Causse, M., & Chounet, A. (2016). Stress-drop variability of shallow earthquakes extracted from a global database of source time functions. *Seismological Research Letters*, 87(4), 912–918. <https://doi.org/10.1785/0220150283>
- Crotwell, H. P., Owens, T. J., & Ritsema, J. (1999). The TauP toolkit: Flexible seismic travel-time and ray-path utilities. *Seismological Research Letters*, 70(2), 154–160. <https://doi.org/10.1785/gssrl.70.2.154>
- Danré, P., Yin, J., Lipovsky, B. P., & Denolle, M. A. (2019). Earthquakes within earthquakes: Patterns in rupture complexity. *Geophysical Research Letters*, 46(13), 7352–7360. <https://doi.org/10.1029/2019GL083093>
- Dascher-Cousineau, K., Brodsky, E. E., Lay, T., & Goebel, T. H. W. (2020). What controls variations in aftershock productivity? *Journal of Geophysical Research: Solid Earth*, 125(2), e2019JB018111. <https://doi.org/10.1029/2019JB018111>
- Denolle, M. A., & Shearer, P. M. (2016). New perspectives on self-similarity for shallow thrust earthquakes. *Journal of Geophysical Research: Solid Earth*, 121(9), 6533–6565. <https://doi.org/10.1002/2016JB013105>
- Díaz-Mojica, J., Cruz-Atienza, V. M., Madariaga, R., Singh, S. K., Tago, J., & Iglesias, A. (2014). Dynamic source inversion of the M6.5 intermediate-depth Zumpango earthquake in central Mexico: A parallel genetic algorithm. *Journal of Geophysical Research: Solid Earth*, 119(10), 7768–7785. <https://doi.org/10.1002/2013JB010854>
- Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3), 425–455. <https://doi.org/10.1093/biomet/81.3.425>
- Douglas, A. (1997). Bandpass filtering to reduce noise on seismograms: Is there a better way? *Bulletin of the Seismological Society of America*, 87(3), 770–777. <https://doi.org/10.1785/BSSA0870030770>
- Duputel, Z., Kanamori, H., Tsai, V. C., Rivera, L., Meng, L., Ampuero, J.-P., & Stock, J. M. (2012). The 2012 Sumatra great earthquake sequence. *Earth and Planetary Science Letters*, 351–352, 247–257. <https://doi.org/10.1016/j.epsl.2012.07.017>
- Dziewonski, A. M., & Anderson, D. L. (1981). Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, 25(4), 297–356. [https://doi.org/10.1016/0031-9201\(81\)90046-7](https://doi.org/10.1016/0031-9201(81)90046-7)
- Eulenfeld, T., Dahm, T., Heimann, S., & Wegler, U. (2021). Fast and robust earthquake source spectra and moment magnitudes from envelope inversion. *Bulletin of the Seismological Society of America*, 112(2), 878–893. <https://doi.org/10.1785/0120210200>
- Frohlich, C. (1989). The nature of deep-focus earthquakes. *Annual Review of Earth and Planetary Sciences*, 17(1), 227–254. <https://doi.org/10.1146/annurev.earth.17.050189.001303>
- Frohlich, C. (2006). *Deep earthquakes*. Cambridge University Press.
- Green, H. W., & Houston, H. (1995). The mechanics of deep earthquakes. *Annual Review of Earth and Planetary Sciences*, 23(1), 169–213. <https://doi.org/10.1146/annurev.earth.23.050195.001125>
- Gutenberg, B., & Richter, C. F. (1949). *Seismicity of the Earth and associated phenomena*. Princeton University Press. Retrieved from <http://archive.org/details/seismicityoftheearth009299mbp>
- Hacker, B. R., Peacock, S. M., Abers, G. A., & Holloway, S. D. (2003). Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *Journal of Geophysical Research*, 108(B1), 2030. <https://doi.org/10.1029/2001JB001129>
- Houston, H., Benz, H. M., & Vidale, J. E. (1998). Time functions of deep earthquakes from broadband and short-period stacks. *Journal of Geophysical Research*, 103(B12), 29895–29913. <https://doi.org/10.1029/98JB02135>
- Huang, Y., Beroza, G. C., & Ellsworth, W. L. (2016). Stress drop estimates of potentially induced earthquakes in the Guy-Greenbrier sequence. *Journal of Geophysical Research: Solid Earth*, 121(9), 6597–6607. <https://doi.org/10.1002/2016JB013067>
- Ide, S., & Beroza, G. C. (2001). Does apparent stress vary with earthquake size? *Geophysical Research Letters*, 28(17), 3349–3352. <https://doi.org/10.1029/2001GL013106>
- Ihmlé, P. F. (1998). On the interpretation of subevents in teleseismic waveforms: The 1994 Bolivia deep earthquake revisited. *Journal of Geophysical Research*, 103(B8), 17919–17932. <https://doi.org/10.1029/98JB00603>
- International Seismological Centre, I. (2022). ISC-GEM earthquake catalogue. *ISC Bulletin*. <https://doi.org/10.31905/d808b825>
- Kanamori, H. (2004). Static and dynamic scaling relations for earthquakes and their implications for rupture speed and stress drop. *Bulletin of the Seismological Society of America*, 94(1), 314–319. <https://doi.org/10.1785/0120030159>

- Kanamori, H., & Brodsky, E. E. (2004). The physics of earthquakes. *Reports on Progress in Physics*, 67(8), 1429–1496. <https://doi.org/10.1088/0034-4885/67/8/R03>
- Kanamori, H., & Rivera, L. (2006). Energy partitioning during an earthquake. *Geophysical Monograph Series*, 170, 3–13. <https://doi.org/10.1029/170GM03>
- Kaneko, Y., & Shearer, P. M. (2015). Variability of seismic source spectra, estimated stress drop, and radiated energy, derived from cohesive-zone models of symmetrical and asymmetrical circular and elliptical ruptures. *Journal of Geophysical Research: Solid Earth*, 120(2), 1053–1079. <https://doi.org/10.1002/2014JB011642>
- Kennet, B. L. N. (1991). Iaspei 1991 seismological tables. *Terra Nova*, 3(2), 122. <https://doi.org/10.1111/j.1365-3121.1991.tb00863.x>
- Kennett, B. L. N., Engdahl, E. R., & Buland, R. (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, 122(1), 108–124. <https://doi.org/10.1111/j.1365-246X.1995.tb03540.x>
- Kikuchi, M., & Fukao, Y. (1987). Inversion of long-period P-waves from great earthquakes along subduction zones. *Tectonophysics*, 144(1), 231–247. [https://doi.org/10.1016/0040-1951\(87\)90020-5](https://doi.org/10.1016/0040-1951(87)90020-5)
- Kirby, S. H., Stein, S., Okal, E. A., & Rubie, D. C. (1996). Metastable mantle phase transformations and deep earthquakes in subducting oceanic lithosphere. *Reviews of Geophysics*, 34(2), 261–306. <https://doi.org/10.1029/96RG01050>
- Kita, S., & Katsumata, K. (2015). Stress drops for intermediate-depth intraslab earthquakes beneath Hokkaido, northern Japan: Differences between the subducting oceanic crust and mantle events. *Geochemistry, Geophysics, Geosystems*, 16(2), 552–562. <https://doi.org/10.1002/2014GC005603>
- Knopoff, L., & Randall, M. J. (1970). The compensated linear-vector dipole: A possible mechanism for deep earthquakes. *Journal of Geophysical Research*, 75(26), 4957–4963. <https://doi.org/10.1029/JB075i026p04957>
- Ko, J. Y.-T., & Kuo, B.-Y. (2016). Low radiation efficiency of the intermediate-depth earthquakes in the Japan subduction zone. *Geophysical Research Letters*, 43(22), 11611–11619. <https://doi.org/10.1002/2016GL070993>
- Lambert, V., Lapusta, N., & Perry, S. (2021). Propagation of large earthquakes as self-healing pulses or mild cracks. *Nature*, 591(7849), 252–258. <https://doi.org/10.1038/s41586-021-03248-1>
- Li, J., Zheng, Y., Thomsen, L., Lapen, T. J., & Fang, X. (2018). Deep earthquakes in subducting slabs hosted in highly anisotropic rock fabric. *Nature Geoscience*, 11(9), 696–700. <https://doi.org/10.1038/s41561-018-0188-3>
- Luo, H., Zeng, H., Shi, Q., Wang, T., Liao, M., Hu, J., & Wei, S. (2023). Could thermal pressurization have induced the frequency-dependent rupture during the 2019 Mw8.0 Peru intermediate-depth earthquake? *Geophysical Journal International*, 232(1), 115–127. <https://doi.org/10.1093/gji/ggac329>
- Madariaga, R. (1976). Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America*, 66(3), 639–666. <https://doi.org/10.1785/BSSA0660030639>
- Marguin, V., & Simpson, G. (2023). Influence of fluids on earthquakes based on numerical modeling. *Journal of Geophysical Research: Solid Earth*, 128(2), e2022JB025132. <https://doi.org/10.1029/2022JB025132>
- Mirwald, A., Cruz-Atienza, V. M., Díaz-Mojica, J., Iglesias, A., Singh, S. K., Villafuerte, C., & Tago, J. (2019). The 19 September 2017 (Mw7.1) intermediate-depth Mexican earthquake: A slow and energetically inefficient deadly shock. *Geophysical Research Letters*, 46(4), 2054–2064. <https://doi.org/10.1029/2018GL080904>
- Montagner, J.-P., & Kennett, B. L. N. (1996). How to reconcile body-wave and normal-mode reference Earth models. *Geophysical Journal International*, 125(1), 229–248. <https://doi.org/10.1111/j.1365-246X.1996.tb06548.x>
- Mousavi, S. M., & Langston, C. A. (2017). Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchro-squeezed domain and its application on earthquake data. *Geophysics*, 82(4), V211–V227. <https://doi.org/10.1190/geo2016-0433.1>
- Neely, J. S., Stein, S., & Spencer, B. D. (2020). Large uncertainties in earthquake stress-drop estimates and their tectonic consequences. *Seismological Research Letters*, 91(4), 2320–2329. <https://doi.org/10.1785/0220200004>
- Nishitsuiji, Y., & Mori, J. (2014). Source parameters and radiation efficiency for intermediate-depth earthquakes in Northeast Japan. *Geophysical Journal International*, 196(2), 1247–1259. <https://doi.org/10.1093/gji/ggt458>
- Noda, H., & Lapusta, N. (2013). Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature*, 493(7433), 518–521. <https://doi.org/10.1038/nature11703>
- Novoselov, A., Balazs, P., & Bokelmann, G. (2022). SEDENOS: SEparating and DENOising seismic signals with dual-path recurrent neural network architecture. *Journal of Geophysical Research: Solid Earth*, 127(3), e2021JB023183. <https://doi.org/10.1029/2021JB023183>
- Oth, A., Parolai, S., Bindi, D., & Wenzel, F. (2009). Source spectra and site response from S waves of intermediate-depth Vrancea, Romania, earthquakes. *Bulletin of the Seismological Society of America*, 99(1), 235–254. <https://doi.org/10.1785/0120080059>
- Park, S., Avouac, J.-P., Zhan, Z., & Gualandri, A. (2023). Weak upper-mantle base revealed by postseismic deformation of a deep earthquake. *Nature*, 615(7952), 1–6. <https://doi.org/10.1038/s41586-022-05689-8>
- Park, S., & Ishii, M. (2015). Inversion for rupture properties based upon 3-D directivity effect and application to deep earthquakes in the Sea of Okhotsk region. *Geophysical Journal International*, 203(2), 1011–1025. <https://doi.org/10.1093/gji/ggv352>
- Pearson, D. G., Brenker, F. E., Nestola, F., McNeill, J., Nasdala, L., Hutchison, M. T., et al. (2014). Hydrous mantle transition zone indicated by ringwoodite included within diamond. *Nature*, 507(7491), 221–224. <https://doi.org/10.1038/nature13080>
- Plümpert, O., John, T., Podladchikov, Y. Y., Vrijmoed, J. C., & Scambelluri, M. (2017). Fluid escape from subduction zones controlled by channel-forming reactive porosity. *Nature Geoscience*, 10(2), 150–156. <https://doi.org/10.1038/ngeo2865>
- Poli, P., & Prieto, G. (2014). Global and along-strike variations of source duration and scaling for intermediate-depth and deep-focus earthquakes. *Geophysical Research Letters*, 41(23), 8315–8324. <https://doi.org/10.1002/2014GL061916>
- Poli, P., & Prieto, G. A. (2016). Global rupture parameters for deep and intermediate-depth earthquakes. *Journal of Geophysical Research: Solid Earth*, 121(12), 8871–8887. <https://doi.org/10.1002/2016JB013521>
- Prieto, G. A. (2022). The multitaper spectrum analysis package in Python. *Seismological Research Letters*, 93(3), 1922–1929. <https://doi.org/10.1785/0220210332>
- Prieto, G. A., Florez, M., Barrett, S. A., Beroza, G. C., Pedraza, P., Blanco, J. F., & Poveda, E. (2013). Seismic evidence for thermal runaway during intermediate-depth earthquake rupture. *Geophysical Research Letters*, 40(23), 6064–6068. <https://doi.org/10.1002/2013GL058109>
- Prieto, G. A., Froment, B., Yu, C., Poli, P., & Abercrombie, R. (2017). Earthquake rupture below the brittle-ductile transition in continental lithospheric mantle. *Science Advances*, 3(3), e1602642. <https://doi.org/10.1126/sciadv.1602642>
- Prieto, G. A., Parker, R. L., & Vernon, F. L., III. (2009). A Fortran 90 library for multitaper spectrum analysis. *Computers & Geosciences*, 35(8), 1701–1710. <https://doi.org/10.1016/j.cageo.2008.06.007>
- Radulian, M., & Popa, M. (1996). Scaling of source parameters for Vrancea (Romania) intermediate depth earthquakes. *Tectonophysics*, 261(1), 67–81. [https://doi.org/10.1016/0040-1951\(96\)00057-1](https://doi.org/10.1016/0040-1951(96)00057-1)

- Rice, J. R. (2006). Heating and weakening of faults during earthquake slip. *Journal of Geophysical Research*, 111(B5), B05311. <https://doi.org/10.1029/2005JB004006>
- Saad, O. M., & Chen, Y. (2020). Deep denoising autoencoder for seismic random noise attenuation. *Geophysics*, 85(4), V367–V376. <https://doi.org/10.1190/geo2019-0468.1>
- Schmandt, B., Jacobsen, S. D., Becker, T. W., Liu, Z., & Dueker, K. G. (2014). Dehydration melting at the top of the lower mantle. *Science*, 344(6189), 1265–1268. <https://doi.org/10.1126/science.1253358>
- Shearer, P. M., Abercrombie, R. E., Trugman, D. T., & Wang, W. (2019). Comparing EGF methods for estimating corner frequency and stress drop from P wave spectra. *Journal of Geophysical Research: Solid Earth*, 124(4), 3966–3986. <https://doi.org/10.1029/2018JB016957>
- Shi, Q. (2023). TeleseismicDenoiser [Software]. DenoTe. <https://doi.org/10.5281/zenodo.7807794>
- Shi, Q., & Denolle, M. A. (2023). Training data of the waveform denoiser to separate teleseismic earthquake and noise signals [Dataset]. UW Photonic Sensing Facility. <https://doi.org/10.6069/69E7-6667>
- Shi, Q., & Wei, S. (2020). Highly heterogeneous pore fluid pressure enabled rupture of orthogonal faults during the 2019 Ridgecrest Mw7.0 earthquake. *Geophysical Research Letters*, 47(20), e2020GL089827. <https://doi.org/10.1029/2020GL089827>
- Singh, S. K., Pacheco, J. F., Bansal, B. K., Pérez-Campos, X., Dattatrayam, R. S., & Suresh, G. (2004). A source study of the Bhuj, India, earthquake of 26 January 2001 (Mw 7.6). *Bulletin of the Seismological Society of America*, 94(4), 1195–1206. <https://doi.org/10.1785/012003212>
- Sobolev, A. V., Asafov, E. V., Gurenko, A. A., Arndt, N. T., Batanova, V. G., Portnyagin, M. V., et al. (2019). Deep hydrous mantle reservoir provides evidence for crustal recycling before 3.3 billion years ago. *Nature*, 571(7766), 555–559. <https://doi.org/10.1038/s41586-019-1399-5>
- Steblov, G. M., Ekström, G., Kogan, M. G., Freymueller, J. T., Titkov, N. N., Vasilenko, N. F., et al. (2014). First geodetic observations of a deep earthquake: The 2013 Sea of Okhotsk Mw 8.3, 611 km-deep, event. *Geophysical Research Letters*, 41(11), 3826–3832. <https://doi.org/10.1002/2014GL060003>
- Stockwell, R., Mansinha, L., & Lowe, R. (1996). Localization of the complex spectrum: The S transform. In *IEEE Transactions on Signal Processing* (Vol. 44(4), pp. 998–1001). <https://doi.org/10.1109/78.492555>
- Thomson, D. (1982). Spectrum estimation and harmonic analysis. In *Proceedings of the IEEE* (Vol. 70(9), pp. 1055–1096). <https://doi.org/10.1109/PROC.1982.12433>
- Tian, D., Wei, S. S., Wang, W., & Wang, F. (2022). Stress drops of intermediate-depth and deep earthquakes in the Tonga slab. *Journal of Geophysical Research: Solid Earth*, 127(10), e2022JB025109. <https://doi.org/10.1029/2022JB025109>
- Tibi, R., Bock, G., & Wiens, D. A. (2003). Source characteristics of large deep earthquakes: Constraint on the faulting mechanism at great depths. *Journal of Geophysical Research*, 108(B2), 2091. <https://doi.org/10.1029/2002JB001948>
- Trugman, D. T., & Shearer, P. M. (2017). Application of an improved spectral decomposition method to examine earthquake source scaling in Southern California. *Journal of Geophysical Research: Solid Earth*, 122(4), 2890–2910. <https://doi.org/10.1002/2017JB013971>
- Tsai, V. C., Nettles, M., Ekström, G., & Dziewonski, A. M. (2005). Multiple CMT source analysis of the 2004 Sumatra earthquake. *Geophysical Research Letters*, 32(17), L17304. <https://doi.org/10.1029/2005GL023813>
- Tschauner, O., Huang, S., Greenberg, E., Prakapenka, V. B., Ma, C., Rossman, G. R., et al. (2018). Ice-VII inclusions in diamonds: Evidence for aqueous fluid in Earth's deep mantle. *Science*, 359(6380), 1136–1139. <https://doi.org/10.1126/science.aao3030>
- Turner, A. R., Ferreira, A. M. G., Berbellini, A., Brantut, N., Faccenda, M., & Kendall, E. (2022). Across-slab propagation and low stress drops of deep earthquakes in the Kuril subduction zone. *Geophysical Research Letters*, 49(16), e2022GL098402. <https://doi.org/10.1029/2022GL098402>
- Vallée, M. (2013). Source time function properties indicate a strain drop independent of earthquake depth and magnitude. *Nature Communications*, 4(1), 2606. <https://doi.org/10.1038/ncomms3606>
- Venkataraman, A., & Kanamori, H. (2004). Observational constraints on the fracture energy of subduction zone earthquakes. *Journal of Geophysical Research*, 109(B5), B05302. <https://doi.org/10.1029/2003JB002549>
- Viesca, R. C., & Garagash, D. I. (2015). Ubiquitous weakening of faults due to thermal pressurization. *Nature Geoscience*, 8(11), 875–879. <https://doi.org/10.1038/ngeo2554>
- Wang, Y., & Day, S. M. (2017). Seismic source spectral properties of crack-like and pulse-like modes of dynamic rupture. *Journal of Geophysical Research: Solid Earth*, 122(8), 6657–6684. <https://doi.org/10.1002/2017JB014454>
- Warren, L. M., & Shearer, P. M. (2006). Systematic determination of earthquake rupture directivity and fault planes from analysis of long-period P-wave spectra. *Geophysical Journal International*, 164(1), 46–62. <https://doi.org/10.1111/j.1365-246X.2005.02769.x>
- Wei, S., Helmberger, D., Zhan, Z., & Graves, R. (2013). Rupture complexity of the Mw 8.3 sea of Okhotsk earthquake: Rapid triggering of complementary earthquakes? *Geophysical Research Letters*, 40(19), 5034–5039. <https://doi.org/10.1002/grl.50977>
- Wiens, D. A. (2001). Seismological constraints on the mechanism of deep earthquakes: Temperature dependence of deep earthquake source properties. *Physics of the Earth and Planetary Interiors*, 127(1), 145–163. [https://doi.org/10.1016/S0031-9201\(01\)00225-4](https://doi.org/10.1016/S0031-9201(01)00225-4)
- Yamasaki, T., & Seno, T. (2003). Double seismic zone and dehydration embrittlement of the subducting slab. *Journal of Geophysical Research*, 108(B4), 2212. <https://doi.org/10.1029/2002JB001918>
- Ye, L., Lay, T., & Kanamori, H. (2020). Anomalously low aftershock productivity of the 2019 MW 8.0 energetic intermediate-depth faulting beneath Peru. *Earth and Planetary Science Letters*, 549, 116528. <https://doi.org/10.1016/j.epsl.2020.116528>
- Ye, L., Lay, T., Kanamori, H., Zhan, Z., & Duputel, Z. (2016). Diverse rupture processes in the 2015 Peru deep earthquake doublet. *Science Advances*, 2(6), e1600581. <https://doi.org/10.1126/sciadv.1600581>
- Yin, J., Denolle, M. A., & He, B. (2022). A multitask encoder–decoder to separate earthquake and ambient noise signal in seismograms. *Geophysical Journal International*, 231(3), 1806–1822. <https://doi.org/10.1093/gji/ggac290>
- Yin, J., Li, Z., & Denolle, M. A. (2021). Source time function clustering reveals patterns in earthquake dynamics. *Seismological Research Letters*, 92(4), 2343–2353. <https://doi.org/10.1785/0220200403>
- Zhan, Z. (2020). Mechanisms and implications of deep earthquakes. *Annual Review of Earth and Planetary Sciences*, 48(1), 147–174. <https://doi.org/10.1146/annurev-earth-053018-060314>
- Zhan, Z., Helmberger, D. V., Kanamori, H., & Shearer, P. M. (2014). Supershear rupture in a Mw 6.7 aftershock of the 2013 Sea of Okhotsk earthquake. *Science*, 345(6193), 204–207. <https://doi.org/10.1126/science.1252717>
- Zhan, Z., Kanamori, H., Tsai, V. C., Helmberger, D. V., & Wei, S. (2014). Rupture complexity of the 1994 Bolivia and 2013 Sea of Okhotsk deep earthquakes. *Earth and Planetary Science Letters*, 385, 89–96. <https://doi.org/10.1016/j.epsl.2013.10.028>
- Zhang, R. (2019). Making convolutional networks shift-invariant again. arXiv. 10.48550/arXiv.1904.11486.
- Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019). Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), 9476–9488. <https://doi.org/10.1109/TGRS.2019.2926772>

- Zhu, W., Mousavi, S. M., & Beroza, G. C. (2020). Seismic signal augmentation to improve generalization of deep neural networks. In *Advances in geophysics* (Vol. 61, pp. 151–177). Elsevier. <https://doi.org/10.1016/bs.agph.2020.07.003>
- Zollo, A., Orefice, A., & Convertito, V. (2014). Source parameter scaling and radiation efficiency of microearthquakes along the Irpinia fault zone in southern Apennines, Italy. *Journal of Geophysical Research: Solid Earth*, 119(4), 3256–3275. <https://doi.org/10.1002/2013JB010116>