Improved observations of deep earthquake ruptures using machine learning

Qibin Shi[1] **and Marine A. Denolle**[1]

[1]Department of Earth and Space Sciences, University of Washington

Key Points:

- A neural network is used to double the number of earthquakes studied by improving the data quality.
- Denoising teleseismic waves improves the source signature in Mw5-6 events and reduces uncertainties
- Large deep earthquake ruptures are dissipative and compact

Corresponding author: Qibin Shi, `qibins@uw.edu`

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) 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 $M_0 \simeq \tau^{4.16}$, and a resulting strong scaling of stress drop and radiated energy with magnitude ($\sigma \simeq M_0^{0.2}$ and $E_R \simeq M_0^{1.23}$). The median radiation efficiency is 5%, a low value compared to shallow 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) 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 that are not related 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 their stress drop and radiated energy are abnormally larger the bigger they are. 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 (Steblov et al., 2014; Luo et al., 2023; Park et al., 2023), and their extreme remoteness yields poor seismic signals on surface sensors. They occur in the deep portion of 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 of seismicity 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 fairly complicated characteristics of the deep earthquakes (Ye et al., 2016; Knopoff & Randall, 1970). 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). Deep earthquakes' stress drops are larger than shallow earthquakes, mostly due to the increased rigidity (Vallée, 2013). Multiple investigations found a strong magnitude dependence of the stress drop, which may be interpreted as dynamic weakening mechanisms (Radulian & Popa, 1996; Oth et al., 2009; Prieto et al., 2013; Poli & Prieto, 2016). Deep earthquakes follow Gutenberg-Richter law (B. Gutenberg & C. F. 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: i) mineral dehydration from metamorphosis processes that release fluids and lubricate faults (i.e., dehydration embrittlement), ii) phase transformation that changes mineral density and volume, and iii) thermal runaway that lowers fault friction from shear heating. The fluids released by mineral dehydration are thought to explain the double-seismic zone (DSZ) (Brudzinski et al., 2007; Hacker et al., 2003; Yamasaki & Seno, 2003; Abers et al., 2013). Whether the released water can penetrate the slab core (Green & Houston, 1995; Boneh et al., 2019) and be transported deeper in the mantle is still under debate (Plümper et al., 2017; Pearson et al., 2014; Schmandt et al., 2014; Tschauner et al., 2018; Sobolev et al., 2019).

Teleseismic observations of deep earthquakes are the most common data available to study these earthquakes. Because small events are more frequent than large earthquakes, moderate-size earthquakes (Mw5-6) 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 (Mw5-6) 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 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 requiring extensive human intervention (Donoho & Johnstone, 1994; Stockwell et al., 1996; Chang et al., 2000; Mousavi & Langston, 2017). 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. 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.

There remain challenges 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) for Mw5-8 earthquakes of the 2000-2021 International Seismological Centre (ISC) earthquake catalog (International 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 would explain intermediate and deep earthquakes.

2 Data Preparation

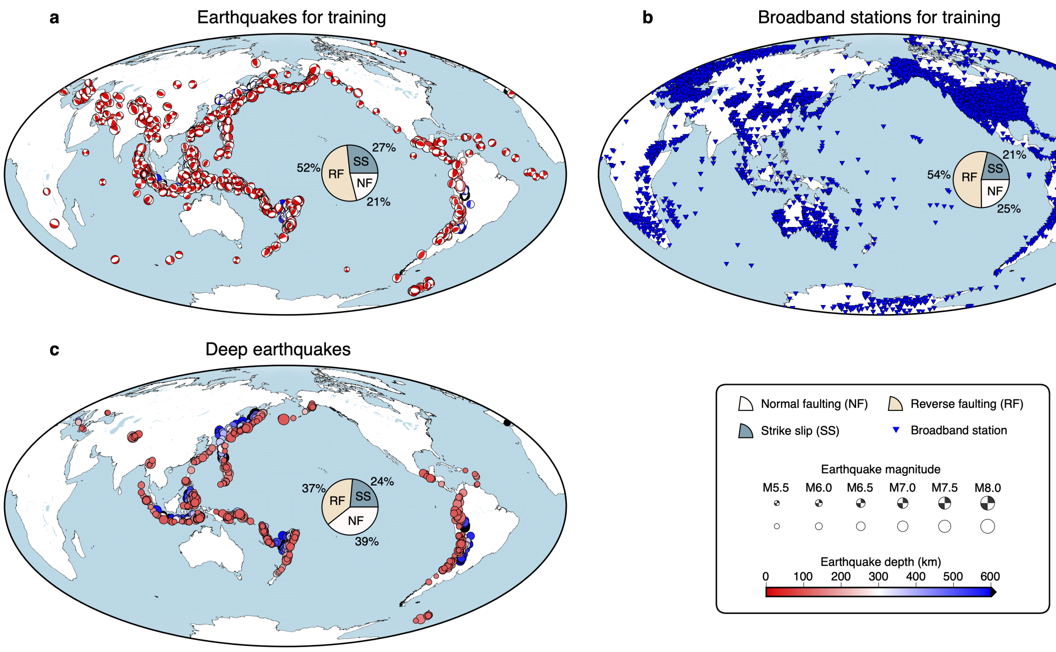

Figure 1. Earthquakes and seismic stations. (a) The 1148 earthquakes with high-SNR recordings used as training data. (b) The FDSN and GSN broadband stations recorded the 45,262 high-SNR teleseismic waveforms of the 1148 earthquakes. (c) The 920 deep earthquakes with low-SNR teleseismic waveforms labeled with focal mechanisms are denoised and tested in this study.

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 Mw5.5+ earthquakes from the 2000-2021 ISC earthquake catalog (International Seismological Centre, 2022) based on focal mechanisms (specifically rake angle) to ensure a relatively even number of strike-slip (306), normal-faulting (242), and reverse-faulting (600) earthquake types. The extracted earthquake list includes events from diverse seismic regions and depths ranging from the surface to 700 km (Figure1a).

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, 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 (Crotwell et al., 1999; Beyreuther et al., 2010) 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 seconds before and 2,500 seconds after the P-arrival. We then calculate the amplitude-based SNR using a noise window (75-10 seconds before) and a signal window (0-75 seconds after the P-wave arrival) with the following definition,

$$SNR = \frac{A_S}{A_N}, \tag{1}$$

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-second 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 WaveDecomNet kernel with feature extraction layers. The stacked neural network on the top encoder branch is a 2-layer convolutional neural network (CNN) with a 1-layer fully connected layer (FCNN) on the optimal training performance. Next, we introduce the architecture of the two-branch encoder-encoder (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 (Zhu et al., 2019; Novoselov et al., 2022), 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 simply duration

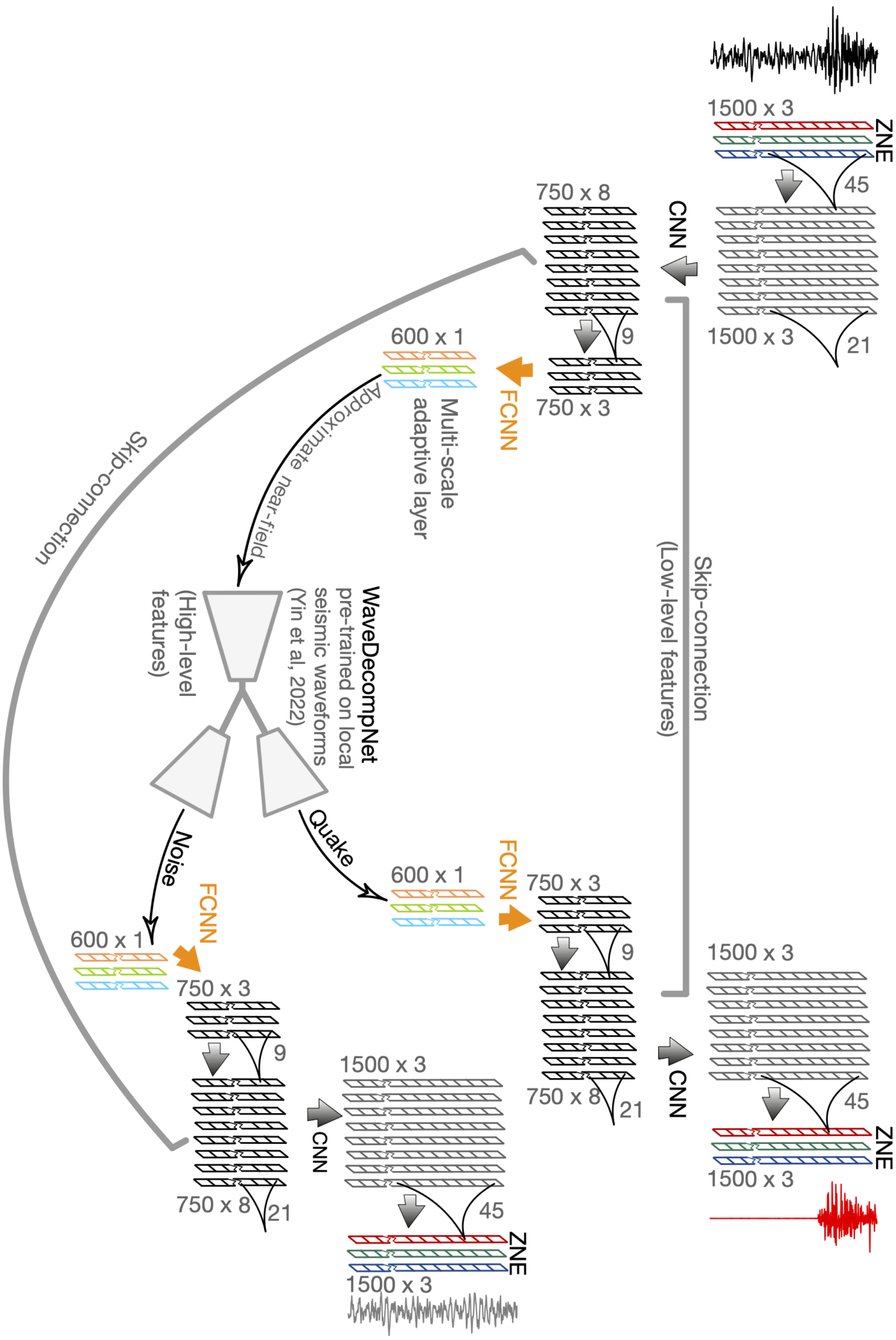

Figure 2. Architecture of the teleseismic wave denoiser, DenoTe. DenoTe is constructed based on the U-net with symmetric structures in the encoding and decoding branches of WaveDecompNet (Yin et al., 2022). The neural network reads composite earthquake waveforms (black) and predicts earthquake (red) and noise (gray) signals through the two output branches, which have the same structure and length. The size, number of channels, and kernel length are indicated for each sub-network. CNN: convolutional neural network. FCNN: fully connected neural network.

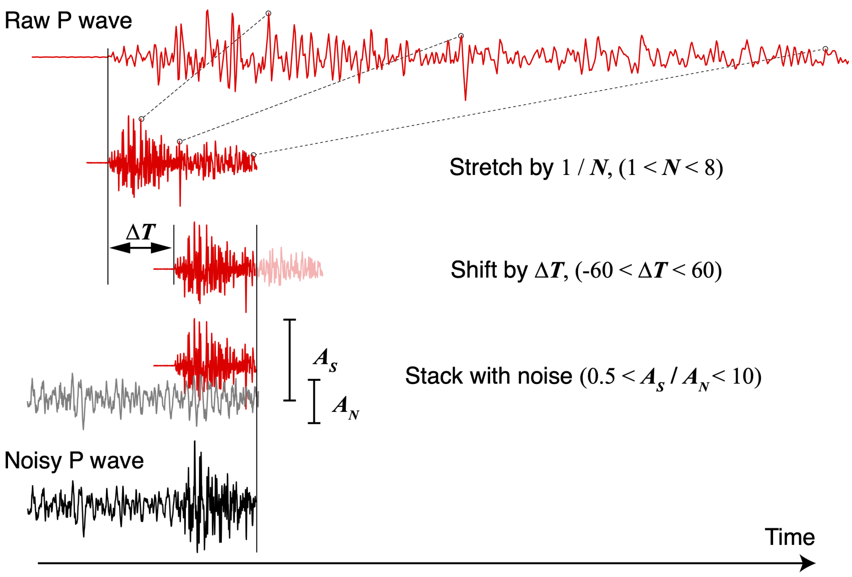

Figure 3. The three steps of data augmentation: the raw high-SNR P wave (red) is 1) stretched, 2) shifted along the time axis, and 3) scaled before it is stacked with the noise (gray) extracted from the same station to compose the noisy waveform (black).

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+), 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 (Mw5-6). 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 seconds. After shifting, we trim the time series to the $-75s \sim +75s$ time window. Thus, the trimmed waveforms mostly include the P wave onsets. In the final augmentation step, we stack each 150-second trace with the 150-second amplified noise extracted from pre-P noise at the 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 of the data is enhanced with each additional training step (epoch), which reduces the possibility of overfitting the training data (Zhu et al., 2020).

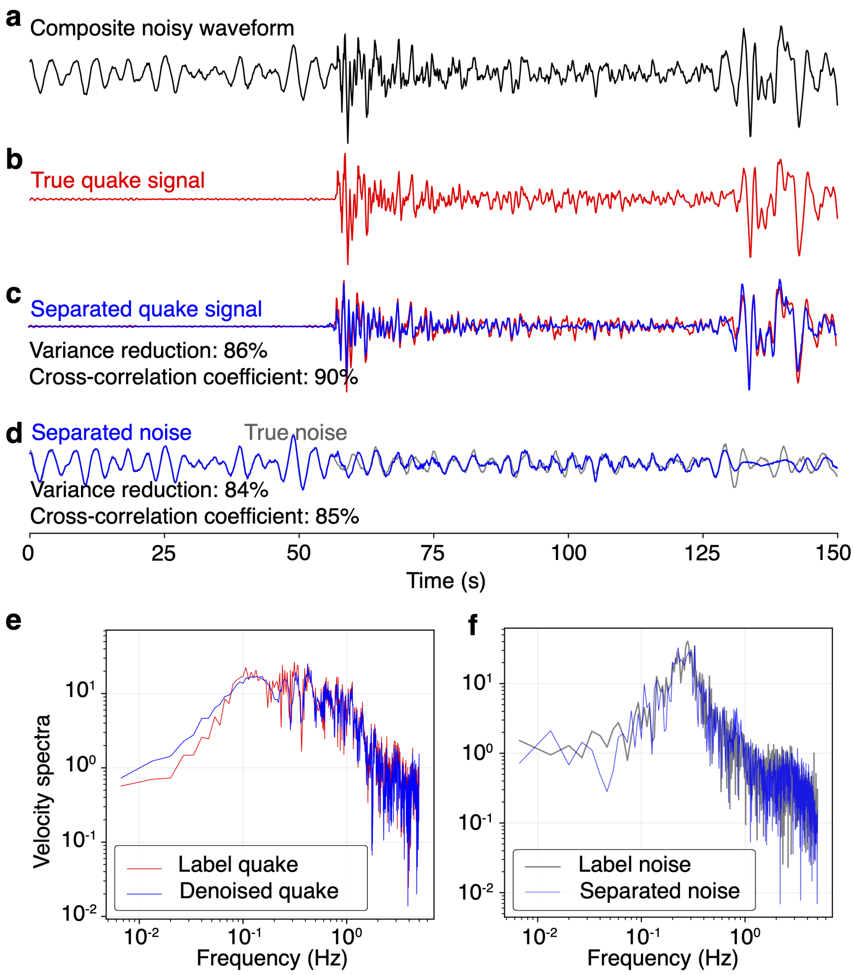

Figure 4. Example of DenoTe's performance. In the time domain: (a) composite waveform, (b) (label) earthquake signal (label data, P-wave, its coda, and the direct S wave), (c) comparison between the labeled (red) and predicted (blue) earthquake signals (and their variance reduction and correlation coefficient),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

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 dataset 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: loss = MSE + 1- CC. The CC is independent of the absolute wave amplitude, typically between -1.0 and 1.0, such that 1-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-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-decent process. This training process is efficient and converges at a low loss of about 0.45 after 140 epochs (see Figure S1). The validation loss computed for every epoch shows closely follows the training loss. The final testing loss is 0.46 (Fig. 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.

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 Figure1c). 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 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 easier to visualize and automatically measure pulse width.

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 extracted from the time domain (source duration and directivity) or the spectral domain (corner frequency, stress drop, radiated energy, and radiation ratio).

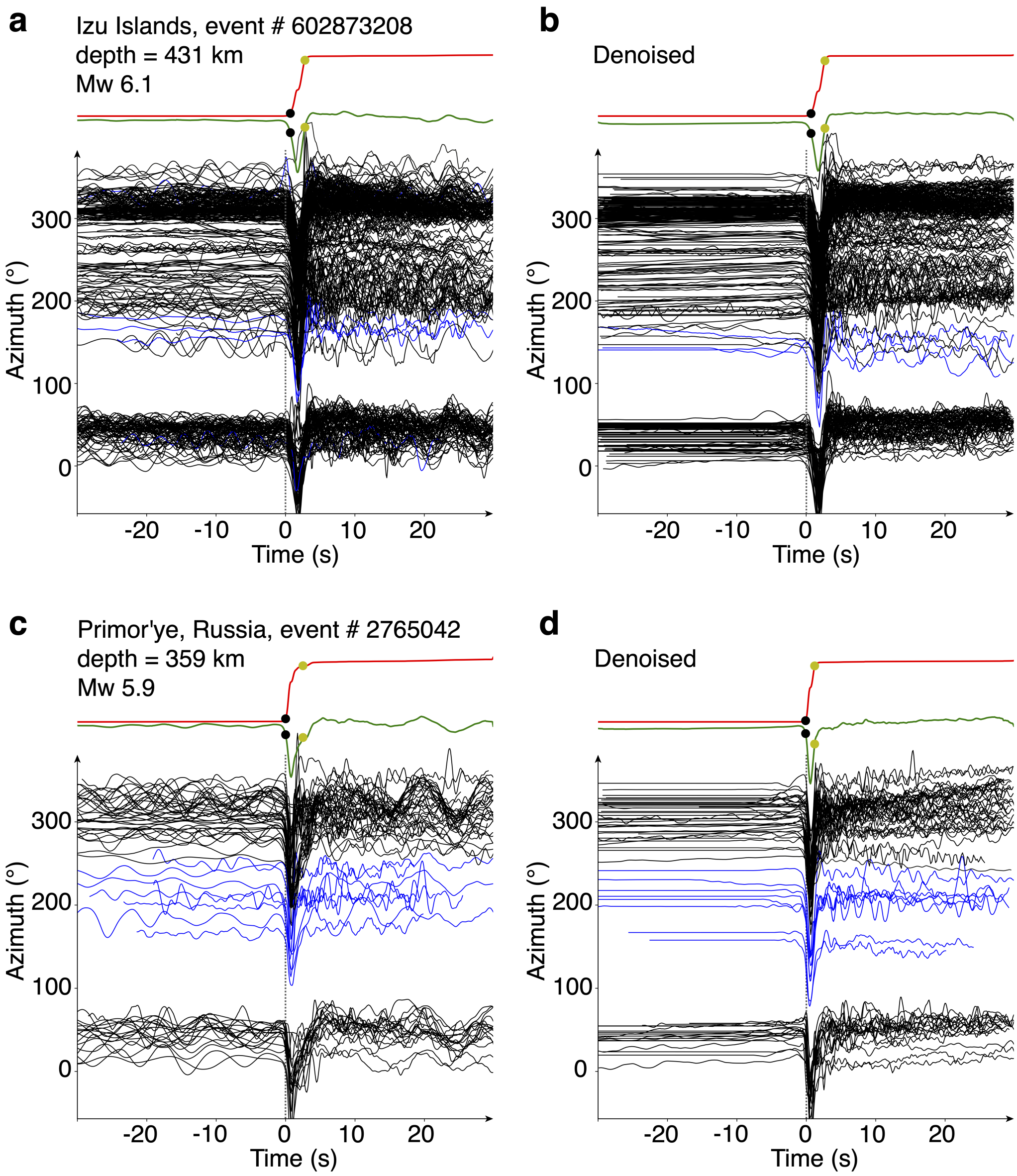

Figure 5. Denoising performance on two representative earthquakes deep earthquakes: the Mw6.1 2013 April 21 earthquake near the Izu Islands in Japan and the Mw5.9 2002 February 1 earthquake at Primor'ye in Russia. (a) and (c) show the original displacement waveforms, and (b) and (d) show the denoised 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.

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 **739** deep Mw5+ 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, 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,

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

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.

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 supplementary 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 (Figure 6a). This increased variability at low magnitudes is typical of studies Allmann and Shearer

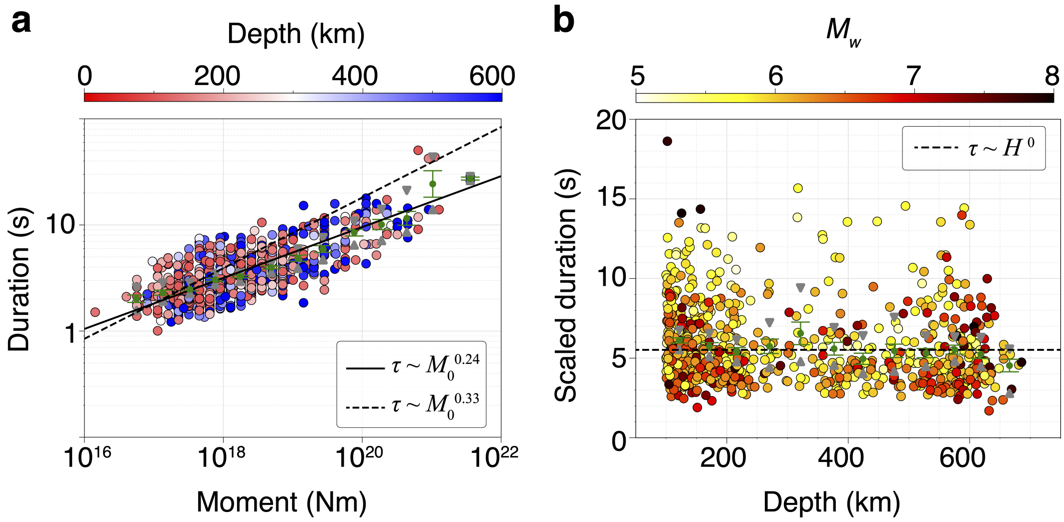

Figure 6. Durations scaling with magnitude and depth. (a) The source duration is shown as a function of the moment with markers color-coded by the event depth and compared with two idealized scaling relationships shown as black lines (the solid line for a scaling of 0.24, the dashed line for a self-similar scaling of 0.33). Each green dot and bar indicate the bootstrapped average of each moment bin and its standard deviation. (b) The magnitude-scaled source duration (eq 2) against depth and color-coded by the event magnitude. The green dots indicate the bootstrapped average and the error bars indicate the standard deviation of the depth bins.

(2009); Denolle and Shearer (2016); Courboulex et al. (2016). As shown in Figure 6a, 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).

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 $a = 0.24$ matches best with the moderate- to large-magnitude earthquakes, and this represents the scaling $\tau \sim M_0^{0.24}$. The measurements of the inverse of corner frequency further confirm the scaling assuming $\tau = 1/f_c$ (see Figure S3). This scaling is similar to what has been found for intermediate and deep earthquakes (Allmann & Shearer, 2009; Turner et al., 2022; Poli & Prieto, 2016).

The depth dependence in scaled duration is well explained by the depth variations in material properties, or equivalently that scaled duration is depth independent. Given a reference magnitude of Mw6.6, 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 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:

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

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 (Fig. 7a), $\cos\theta_i$ is

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

where the angle parameters are explained and illustrated in Figure S4. Each source-station geometry provides a unique set of geometrical parameters. We know ϕ_i and γ_i from 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 τ and 1.4 V_r τ with 8 grid points.

In order 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 draw a three-dimensional distribution of the relative durations because the P-wave rays from the source to receivers have specific take-off angles and azimuths.

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. High Doppler strength 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

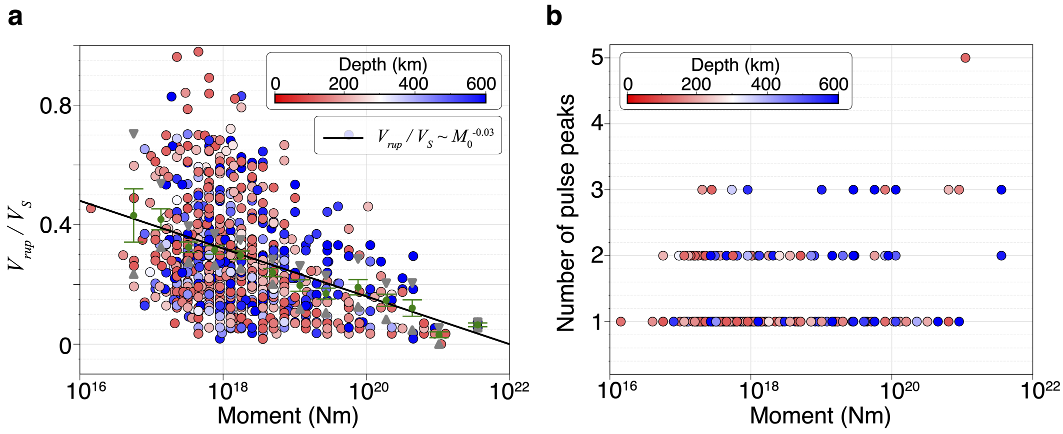

Figure 7. The Doppler effect of deep earthquakes analyzed in this study. (a) The equivalent unilateral rupture speed ratio to the S-wave velocity near the earthquake source is plotted to show the relation with the moment, color-coded by event depth. (b) The number of peaks of the source time function in relation to seismic moment color-coded by event depth.

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 i) the weakening of material beyond the seismogenic width (i.e., the slab) or ii) the growing complexity of the rupture processes, which can be involved with multiple faults or multiple mechanisms during 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, we cannot 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 (Kikuchi & Fukao, 1987; Houston et al., 1998; Ihmlé, 1998; Antolik et al., 1999; Tibi et al., 2003; Tsai et al., 2005; Duputel et al., 2012; Wei et al., 2013; Zhan, Kanamori, et al., 2014; Danré et al., 2019; Shi & Wei, 2020; Yin et al., 2021). 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 7c. The waveform resolution (<4 Hz) is sufficient for Mw > 6 events and well below some Mw5.0-6.0 earthquakes. Three subevents are only detected for Mw > 5.5, and smaller events present fewer subevents (i.e., 1 or 2) as shown in Figure 7b. Larger earthquakes have a few but more subevents, but overall, deep earthquakes are simpler ruptures with fewer subevents confirming Yin et al. (2021) and the hypothesis that deep earthquakes are rather crack-like.

4.4 Spectral Fitting

The far-field P wave displacement waveforms are an approximation to the moment-rate function. Their amplitudes are controlled by radiation patterns and geometrical spreading, which are 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,

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

where $t^* = 0.3$ for the P waves that originate from the mantle (Poli & Prieto, 2016). 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).

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

where the two parameters to find are the falloff 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 between 0 and 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 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. The difference in spectral shapes between the synthetic and stacked spectra is reduced after denoising.

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 vi-

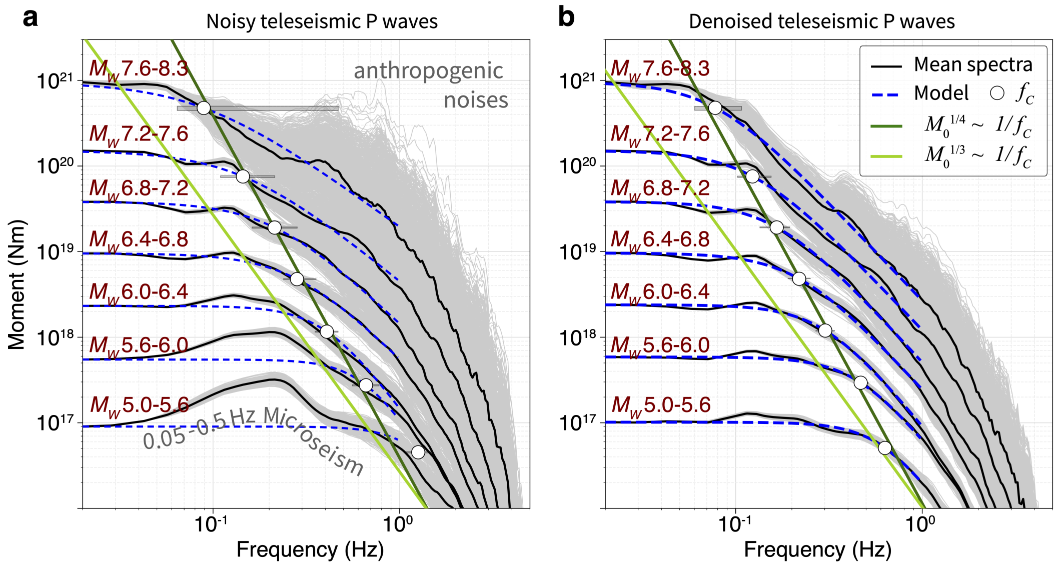

Figure 8. Spectra averaged in magnitude bins. (a) The noisy spectra are divided into seven magnitude groups, as indicated on the left, and bootstrapped in each group 1000 times to compute the average spectra (gray). The median of the bootstrapped spectra mean (black lines) is well fit by the spectral model (blue dashed lines) after searching for the optimal corner frequency (yellow dots) and high-frequency fall-off rate. The corner frequency is marked as white circles with uncertainties shown as gray bars. (b) Same as (a) for the denoised waveforms.

sual correlation that $M_0 \propto f_c^{-4}$, again supporting a deviation from a self-similar behavior. This result holds when considering the 739 individual estimates of f_c (Figure S3) and confirms the inverse relation between duration τ and moment, $M_0 \sim \tau^{4.17}$, illustrated in Figure 6a.

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 retrieval of corner frequency for magnitude Mw 5-6.5. 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

Since the spectra are well fit using a single-corner frequency model and the weak directivity effects, we propose using a circular crack model of rupture for deep earthquakes. 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 (Brune, 1970) later updated by (Madariaga, 1976) to relate event duration and moment to stress drop $\Delta\sigma$:

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

where the geometrical parameter 7/16 is used for a circular crack, the radius of the crack is estimated as $0.35V_S/f_c$. 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 9. 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

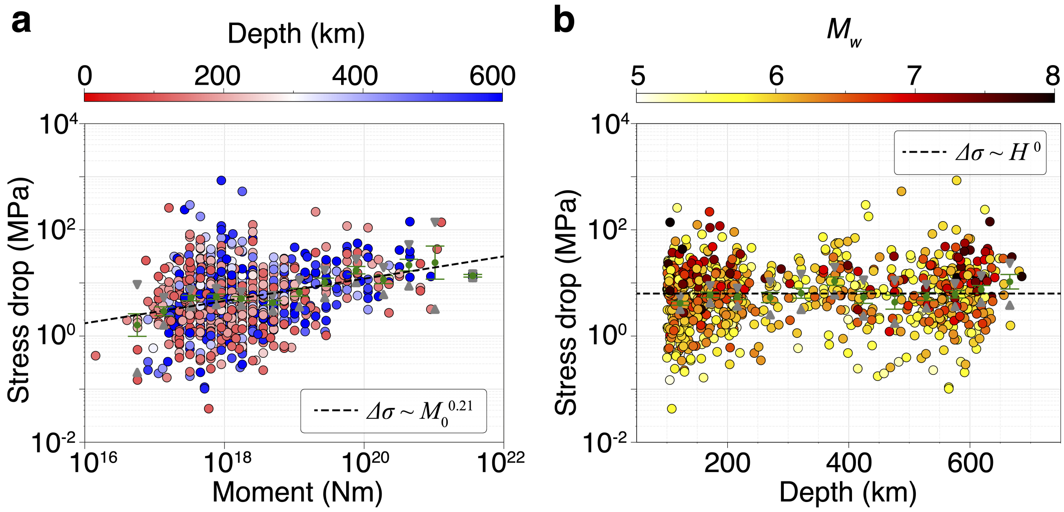

Figure 9. Stress Drop, Depth, and Magnitude. (a) The stress drop is shown against moment, color-coded by the event depth, with the bootstrapped mean of each magnitude bin shown in green and the best-fit scaling relationship denoted by the dashed line. (b) The stress drop is shown against depth, color-coded by the event magnitude. The bootstrapped mean on each depth bin is shown in green, and the best-fit scaling relationship is denoted by the dashed line.

and find the exponent $a = 0.2$. The resulting strong scaling suggests that if the Mw 5.0 earthquakes have a stress drop of about 1.8 MPa, the Mw 7.5 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 (Figure 9). We bootstrap the stress drop 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, such as $\Delta\sigma \sim M_0^{0.21}$. 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.

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 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).

We do not see any strong dependence between stress drop and depth (Fig. 9). We measure an increased variability of the shallowest intermediate-depth earthquakes, which may indicate that we have less 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. 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 broad range of frequencies, to which radiated energy is particularly sensitive. We estimate the radiated P-wave energy using,

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

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 $\hat{S}(f)$, α is the P wave velocity at the location of the source. The shear modulus μ is calculated with the shear-wave velocity of the PREM model, and the seismic moment M_0 is calculated from moment-magnitude.

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

$$\sigma_a = \mu E_R / M_0, \tag{9}$$

In general, the observed spectra well match the model $\hat{S}'(f)$ in equation 6 within 0-1 Hz (see Figure 8). Higher than 1 Hz, the observed spectra have a steeper fall-off than the model, which implies that attenuation may be frequency dependent and under-corrected at higher frequencies. Ide and Beroza (2001) has indicated that the source spectrum at frequencies higher than ten 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 over 0-1 Hz and modeled spectra integrated over 1-4 Hz.

Similar to (Boatwright & Choy, 1986; Convers & Newman, 2011; Poli & Prieto, 2016; Denolle & Shearer, 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 (Knopoff & Randall, 1970; Frohlich, 1989; Green & Houston, 1995), 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 1.23. 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 (Venkataraman & Kanamori, 2004; Baltay et al., 2010; Convers & Newman, 2011), though Denolle and Shearer (2016) found this was true regardless of the fault geometry.

We show the moment-dependent radiated energy derived from the noisy and denoised P waves in Figure 10a and b, respectively. Similar to the other measurements, denoising reduces the variability of the radiated energy measurements but does not alter the general trend of the scaling (Figure S9).

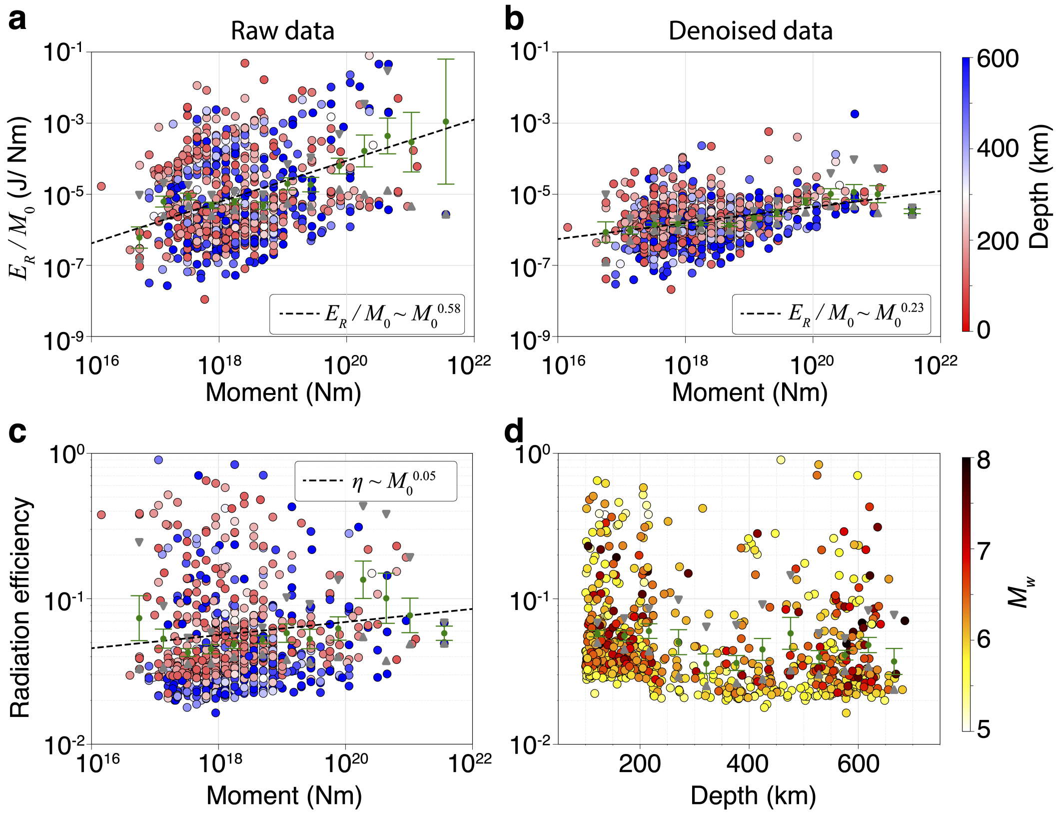

Figure 10. Radiation energy and efficiency of deep earthquakes. Radiated energy as a function of seismic moment and color-coded with depth. The green dots and bars indicate the average logarithmic radiated energy bootstrapped in magnitude bins and the best-fitting regression coefficients, respectively, for the raw, attenuation-corrected waveforms (a) and after denoising, attenuation-corrected waveforms (b). (c) Radiation efficiency against moment as markers color-coded with event depth with the binned average efficiency. (d) Radiation efficiency against depth color-coded by event magnitude, with green dots denoting the bootstrapped average efficiency in each fine depth bin.

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:

$$\eta_R = \frac{2\mu E_R}{\Delta\sigma M_0}, \tag{10}$$

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 (Poli & Prieto, 2016; Prieto et al., 2013; Wiens, 2001). These values are typically much lower than those reported for crustal earthquakes (Venkataraman & Kanamori, 2004; Singh et al., 2004; Zollo et al., 2014; Prieto et al., 2017; Lambert et al., 2021). 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 also observe a weak moment-dependence of radiation efficiency (Figure 10c), 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 Figure 10d). 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 deep earthquake signals relative to shallower earthquake signals, which would give an apparent higher radiated energy. Because radiation efficiency as calculated in equation 10 is effectively proportional to V_P^5/V_S^5, 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 AK135f velocity model (Kennett et al., 1995; Montagner & Kennett, 1996) see Figure S10 for comparison.

4.8 Fracture Energy

Fracture energy is the energy spent to create the fracture. 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:

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

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, circular model as $S = M_0/[\mu\pi(0.35V_S\tau)^2]$. We use τ as our time-domain duration estimate in this example. 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 recover to higher friction when the slip stops (Viesca & Garagash, 2015). We show the estimated values in Figure 11.

In general, deep earthquakes exhibit slightly higher fracture energy, discussed earlier, with a slightly lower radiation efficiency. But overall, both intermediate-depth and

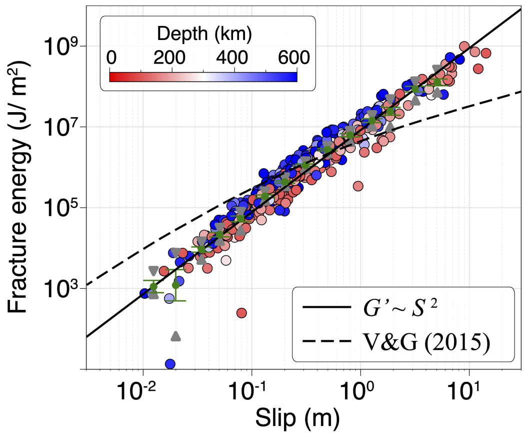

Figure 11. Fracture energy against average slip inferred from seismic observations, color-coded with depth. A linear regression of slip is fitted to the bootstrapped mean values of the binned data, and the best slopes are found for the earthquakes with estimates of average slip as shown in black solid line, in contrast to the power-law scaling by Viesca and Garagash (2015).

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 $G' \sim S^2$ is overall satisfied with our observations. 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 (Noda & Lapusta, 2013; Marguin & Simpson, 2023). 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 imply that the fault weakening is likely to be persistent during the slip growth so that fracture energy keeps 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 Mw>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 Mw8.3 Bolivian earthquake. Park and Ishii (2015) derived the average rupture speed for the 2012 Mw7.7 (2.7 km/s, 583 km) and 2013 Mw8.3 (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 Mw4.8 Wyoming earthquake (75 km). Díaz-Mojica et al. (2014) used an elliptical patch approach to study the 2011 Mw6.5 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 Mw7.1 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 Mw6.7 Sea of Okhotsk earthquake (642 km), implying a very different rupture process relative to the nearby 2013 Mw8.3 Okhotsk Earthquake. This may be confirmed by the larger variability of Doppler ratios we find for Mw5.0-6.9 earthquakes.

The moderate-magnitude earthquakes ($10^{16} < M_0 < 10^{19}$ N m) 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 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 ($M_0 \sim f_c^{-4}$) 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 (Prieto et al., 2013; Poli & Prieto, 2016). We obtain a similar moment-scaling of stress drop $\Delta\sigma \sim M_0^{0.21}$ for Mw5-8 earthquakes at a 100-700 km depth range. This contrasts with shallow earthquakes, where stress drop tends to be scale-invariant (Allmann & Shearer, 2009; Denolle & Shearer, 2016; Courboulex et al., 2016). 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 Figure 9) for a broad range of seismic moment ($-8 < MW < 9$), and reported no evident scaling of stress drop with earthquake size.

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.

In spite of 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 (Fig. 11). 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 y Zhan (2020) is practical 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.

In general, deep earthquakes have relatively simple rupture processes compared to crustal earthquakes because of the fewer subevents identified from their source time functions. This feature may favor that deep earthquakes tend to start on the faults with preferred orientation (e.g., along the metastable olivine wedge or along the pre-existing intraplate 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 Mw5-5.5 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 Figure S11. 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 in the smaller earthquake magnitudes. 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 that have found a strong scaling of stress drop and scaled energy with earthquake magnitude, which suggests weakening mechanisms stronger with earthquake size.

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. In general, we find that typical stress drops of 1-10 MPa and low scaled energy (10^{-5}), 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.

7 Open Research

The software package for denoising is developed using PyTorch. It is named "DenoTe" and can be accessed from `https://github.com/qibinshi/TeleseismicDenoiser`. We use data from the 1078 networks of the FDSN archive. The digital object identifier (DOI) of all 1078 networks can be found in the supplementary materials. The minimally pre-processed seismic data used for training the neural network can be accessed at `https://dasway.ess.washington.edu/qibins/Psnr25_lp4_2000-2021.hdf5` and the waveform data and metadata for the deep eearthquake analysis can be accessed at `http://dasway.ess.washington.edu/qibins/deepquake_M5.5_6_data_metadata.zip`. The earthquake catalog for selecting the waveform data is downloaded from ISC `http://www.isc.ac.uk/`.

Acknowledgments

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. Conceptualization: QS, MD. Data curation: QS. Formal Analysis: QS. Funding acquisition: MD. Investigation QS, MD. Methodology: QS, MD. Project administration: MD, QS. Resources: MD. Software: QS. Supervision: MD. Validation: QS, Yiyu Ni. Visualization: QS. Writing – original draft: QS. Writing – review & editing: QS, MD. 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).

References

Abercrombie, R. E., & Rice, J. R. (2005, August). Can observations of earthquake scaling constrain slip weakening? *Geophysical Journal International*, *162*(2), 406–424. Retrieved 2022-12-04, from `https://doi.org/10.1111/j.1365-246X.2005.02579.x` doi: 10.1111/j.1365-246X.2005.02579.x

Abers, G. A., Nakajima, J., van Keken, P. E., Kita, S., & Hacker, B. R. (2013, May). Thermal–petrological controls on the location of earthquakes within subducting plates. *Earth and Planetary Science Letters*, *369-370*, 178–187. Retrieved 2023-02-23, from `https://www.sciencedirect.com/science/article/pii/S0012821X1300143X` doi: 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: Solid Earth*, *114*(B1). Retrieved 2023-05-01, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2008JB005821` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2008JB005821) doi: 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: Solid Earth*, *104*(B1), 863–894. Retrieved 2023-05-31, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/1998JB900042` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/1998JB900042) doi: 10.1029/1998JB900042

B. Gutenberg, & C. F. Richter. (1949). *Seismicity Of The Earth And Associated Phenomena.* Princeton University Press. Retrieved 2023-01-31, from `http://archive.org/details/seismicityofthee009299mbp`

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: Solid Earth*, *115*(B8). Retrieved 2023-05-31, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2009JB006736` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009JB006736) doi: 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. Retrieved 2023-06-22, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/95GL01089` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/95GL01089) doi: 10.1029/95GL01089

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010, May). ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, *81*(3), 530–533. Retrieved 2023-04-28, from `https://doi.org/10.1785/gssrl.81.3.530` doi: 10.1785/gssrl.81.3.530

Boatwright, J., & Choy, G. L. (1986). Teleseismic estimates of the energy radiated by shallow earthquakes. *Journal of Geophysical Research*, *91*(B2), 2095. Retrieved 2022-08-24, from `http://doi.wiley.com/10.1029/JB091iB02p02095` doi: 10.1029/JB091iB02p02095

Boneh, Y., Schottenfels, E., Kwong, K., van Zelst, I., Tong, X., Eimer, M., ... Zhan, Z. (2019). Intermediate-Depth Earthquakes Controlled by Incoming Plate Hydration Along Bending-Related Faults. *Geophysical Research Letters*, *46*(7), 3688–3697. Retrieved 2023-02-23, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL081585 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL081585) doi: 10.1029/2018GL081585

Brudzinski, M. R., Thurber, C. H., Hacker, B. R., & Engdahl, E. R. (2007, June). Global Prevalence of Double Benioff Zones. *Science*, *316*(5830), 1472–1474. Retrieved 2023-02-23, from https://www.science.org/doi/full/10.1126/science.1139204 (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1139204

Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of Geophysical Research (1896-1977)*, *75*(26), 4997–5009. Retrieved 2023-02-13, from https://onlinelibrary.wiley.com/doi/abs/10.1029/JB075i026p04997 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/JB075i026p04997) doi: 10.1029/JB075i026p04997

Chang, S., Yu, B., & Vetterli, M. (2000, September). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, *9*(9), 1532–1546. (Conference Name: IEEE Transactions on Image Processing) doi: 10.1109/83.862633

Cocco, M., Tinti, E., & Cirella, A. (2016, October). On the scale dependence of earthquake stress drop. *Journal of Seismology*, *20*(4), 1151–1170. Retrieved 2023-06-23, from https://doi.org/10.1007/s10950-016-9594-4 doi: 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: Solid Earth*, *116*(B8). Retrieved 2023-05-31, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2010JB007928 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2010JB007928) doi: 10.1029/2010JB007928

Courboulex, F., Vallée, M., Causse, M., & Chounet, A. (2016, May). Stress-Drop Variability of Shallow Earthquakes Extracted from a Global Database of Source Time Functions. *Seismological Research Letters*, *87*(4), 912–918. Retrieved 2023-02-15, from https://doi.org/10.1785/0220150283 doi: 10.1785/0220150283

Crotwell, H. P., Owens, T. J., & Ritsema, J. (1999, March). The TauP Toolkit: Flexible Seismic Travel-time and Ray-path Utilities. *Seismological Research Letters*, *70*(2), 154–160. Retrieved 2023-03-01, from https://doi.org/10.1785/gssrl.70.2.154 doi: 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. Retrieved 2023-05-31, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083093 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL083093) doi: 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. Retrieved 2023-05-18, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JB018111 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019JB018111) doi: 10.1029/2019JB018111

Denolle, M. A., & Shearer, P. M. (2016, September). New perspectives on self-similarity for shallow thrust earthquakes. *Journal of Geophysical Research: Solid Earth*, *121*(9), 6533–6565. Retrieved 2022-09-21, from https://onlinelibrary.wiley.com/doi/10.1002/2016JB013105 doi:

10.1002/2016JB013105

Donoho, D. L., & Johnstone, I. M. (1994, September). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, *81*(3), 425–455. Retrieved 2023-03-09, from https://doi.org/10.1093/biomet/81.3.425 doi: 10.1093/biomet/81.3.425

Douglas, A. (1997, June). Bandpass filtering to reduce noise on seismograms: Is there a better way? *Bulletin of the Seismological Society of America*, *87*(3), 770–777. Retrieved 2023-03-09, from https://doi.org/10.1785/BSSA0870030770 doi: 10.1785/BSSA0870030770

Duputel, Z., Kanamori, H., Tsai, V. C., Rivera, L., Meng, L., Ampuero, J.-P., & Stock, J. M. (2012, October). The 2012 Sumatra great earthquake sequence. *Earth and Planetary Science Letters*, *351-352*, 247–257. Retrieved 2023-05-31, from https://www.sciencedirect.com/science/article/pii/S0012821X12003846 doi: 10.1016/j.epsl.2012.07.017

Dziewonski, A. M., & Anderson, D. L. (1981, June). Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, *25*(4), 297–356. Retrieved 2023-03-12, from https://www.sciencedirect.com/science/article/pii/0031920181900467 doi: 10.1016/0031-9201(81)90046-7

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. Retrieved 2023-06-23, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2013JB010854 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013JB010854) doi: 10.1002/2013JB010854

Frohlich, C. (1989). The Nature of Deep-Focus Earthquakes. *Annual Review of Earth and Planetary Sciences*, *17*(1), 227–254. Retrieved 2023-02-20, from https://doi.org/10.1146/annurev.ea.17.050189.001303 (_eprint: https://doi.org/10.1146/annurev.ea.17.050189.001303) doi: 10.1146/annurev.ea.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. Retrieved 2023-02-20, from https://doi.org/10.1146/annurev.ea.23.050195.001125 (_eprint: https://doi.org/10.1146/annurev.ea.23.050195.001125) doi: 10.1146/annurev.ea.23.050195.001125

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: Solid Earth*, *108*(B1). Retrieved 2023-02-23, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2001JB001129 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2001JB001129) doi: 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: Solid Earth*, *103*(B12), 29895–29913. Retrieved 2023-02-14, from https://onlinelibrary.wiley.com/doi/abs/10.1029/98JB02135 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/98JB02135) doi: 10.1029/98JB02135

Ide, S., & Beroza, G. C. (2001, September). Does apparent stress vary with earthquake size? *Geophysical Research Letters*, *28*(17), 3349–3352. Retrieved 2022-08-31, from http://doi.wiley.com/10.1029/2001GL013106 doi: 10.1029/2001GL013106

Ihmlé, P. F. (1998). On the interpretation of subevents in teleseismic waveforms: The 1994 Bolivia deep earthquake revisited. *Journal of Geophysi-*

cal Research: Solid Earth, *103*(B8), 17919–17932. Retrieved 2023-05-31, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/98JB00603` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/98JB00603) doi: 10.1029/98JB00603

International Seismological Centre, I. (2022). ISC-GEM Earthquake Catalogue. *ISC Bulletin*. doi: https://doi.org/10.31905/d808b825

Kanamori, H. (2004, February). 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. Retrieved 2023-06-23, from `https://pubs.geoscienceworld.org/bssa/article/94/1/314-319/103081` doi: 10.1785/0120030159

Kanamori, H., & Brodsky, E. E. (2004, August). The physics of earthquakes. *Reports on Progress in Physics*, *67*(8), 1429–1496. Retrieved 2023-02-16, from `https://iopscience.iop.org/article/10.1088/0034-4885/67/8/R03` doi: 10.1088/0034-4885/67/8/R03

Kanamori, H., & Rivera, L. (2006). Energy partitioning during an earthquake. In *Geophysical Monograph Series* (Vol. 170, pp. 3–13). Washington, D. C.: American Geophysical Union. Retrieved 2023-05-29, from `https://onlinelibrary.wiley.com/doi/10.1029/170GM03` doi: 10.1029/170GM03

Kennet, B. L. N. (1991). Iaspei 1991 Seismological Tables. *Terra Nova*, *3*(2), 122–122. Retrieved 2023-03-12, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3121.1991.tb00863.x` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-3121.1991.tb00863.x) doi: 10.1111/j.1365-3121.1991.tb00863.x

Kennett, B. L. N., Engdahl, E. R., & Buland, R. (1995, July). Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, *122*(1), 108–124. Retrieved 2023-03-15, from `https://doi.org/10.1111/j.1365-246X.1995.tb03540.x` doi: 10.1111/j.1365-246X.1995.tb03540.x

Kikuchi, M., & Fukao, Y. (1987, December). Inversion of long-period P-waves from great earthquakes along subduction zones. *Tectonophysics*, *144*(1), 231–247. Retrieved 2023-05-31, from `https://www.sciencedirect.com/science/article/pii/0040195187900205` doi: 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. Retrieved 2023-02-20, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/96RG01050` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/96RG01050) doi: 10.1029/96RG01050

Knopoff, L., & Randall, M. J. (1970). The compensated linear-vector dipole: A possible mechanism for deep earthquakes. *Journal of Geophysical Research (1896-1977)*, *75*(26), 4957–4963. Retrieved 2023-01-31, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/JB075i026p04957` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/JB075i026p04957) doi: 10.1029/JB075i026p04957

Lambert, V., Lapusta, N., & Perry, S. (2021, March). Propagation of large earthquakes as self-healing pulses or mild cracks. *Nature*, *591*(7849), 252–258. Retrieved 2023-05-15, from `https://www.nature.com/articles/s41586-021-03248-1` (Number: 7849 Publisher: Nature Publishing Group) doi: 10.1038/s41586-021-03248-1

Li, J., Zheng, Y., Thomsen, L., Lapen, T. J., & Fang, X. (2018, September). Deep earthquakes in subducting slabs hosted in highly anisotropic rock fabric. *Nature Geoscience*, *11*(9), 696–700. Retrieved 2023-03-15, from `https://www.nature.com/articles/s41561-018-0188-3` (Number: 9 Publisher: Nature Publishing Group) doi: 10.1038/s41561-018-0188-3

Luo, H., Zeng, H., Shi, Q., Wang, T., Liao, M., Hu, J., & Wei, S. (2023, January). 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. Retrieved 2023-01-31, from `https://doi.org/10.1093/gji/ggac329` doi: 10.1093/gji/ggac329

Madariaga, R. (1976, June). Dynamics of an expanding circular fault. *Bulletin of the Seismological Society of America*, *66*(3), 639–666. Retrieved 2023-02-15, from `https://doi.org/10.1785/BSSA0660030639` doi: 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. Retrieved 2023-06-02, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2022JB025132` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022JB025132) doi: 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. Retrieved 2023-06-23, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080904` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018GL080904) doi: 10.1029/2018GL080904

Montagner, J.-P., & Kennett, B. L. N. (1996, April). How to reconcile body-wave and normal-mode reference earth models. *Geophysical Journal International*, *125*(1), 229–248. Retrieved 2023-03-15, from `https://doi.org/10.1111/j.1365-246X.1996.tb06548.x` doi: 10.1111/j.1365-246X.1996.tb06548.x

Mousavi, S. M., & Langston, C. A. (2017, May). Automatic noise-removal/signal-removal based on general cross-validation thresholding in synchrosqueezed domain and its application on earthquake data. *Geophysics*, *82*(4), V211–V227. Retrieved 2023-03-09, from `https://doi.org/10.1190/geo2016-0433.1` doi: 10.1190/geo2016-0433.1

Noda, H., & Lapusta, N. (2013, January). Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature*, *493*(7433), 518–521. Retrieved 2023-05-31, from `https://www.nature.com/articles/nature11703` (Number: 7433 Publisher: Nature Publishing Group) doi: 10.1038/nature11703

Novoselov, A., Balazs, P., & Bokelmann, G. (2022, March). SEDENOSS: SEparating and DENOising Seismic Signals With Dual-Path Recurrent Neural Network Architecture. *Journal of Geophysical Research: Solid Earth*, *127*(3). Retrieved 2022-08-11, from `https://onlinelibrary.wiley.com/doi/10.1029/2021JB023183` doi: 10.1029/2021JB023183

Oth, A., Parolai, S., Bindi, D., & Wenzel, F. (2009, February). 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. Retrieved 2023-05-18, from `https://doi.org/10.1785/0120080059` doi: 10.1785/0120080059

Park, S., Avouac, J.-P., Zhan, Z., & Gualandi, A. (2023, February). Weak upper-mantle base revealed by postseismic deformation of a deep earthquake. *Nature*, 1–6. Retrieved 2023-03-01, from `https://www.nature.com/articles/s41586-022-05689-8` (Publisher: Nature Publishing Group) doi: 10.1038/s41586-022-05689-8

Park, S., & Ishii, M. (2015, November). 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. Retrieved 2023-02-03, from `https://doi.org/10.1093/gji/ggv352` doi: 10.1093/gji/ggv352

Pearson, D. G., Brenker, F. E., Nestola, F., McNeill, J., Nasdala, L., Hutchison, M. T., ... Vincze, L. (2014, March). Hydrous mantle transition zone indicated by ringwoodite included within diamond. *Nature*, *507*(7491), 221–224. Retrieved 2023-02-23, from https://www.nature.com/articles/nature13080 (Number: 7491 Publisher: Nature Publishing Group) doi: 10.1038/nature13080

Plümper, O., John, T., Podladchikov, Y. Y., Vrijmoed, J. C., & Scambelluri, M. (2017, February). Fluid escape from subduction zones controlled by channel-forming reactive porosity. *Nature Geoscience*, *10*(2), 150–156. Retrieved 2023-02-23, from https://www.nature.com/articles/ngeo2865 (Number: 2 Publisher: Nature Publishing Group) doi: 10.1038/ngeo2865

Poli, P., & Prieto, G. (2014, December). Global and along-strike variations of source duration and scaling for intermediate-depth and deep-focus earthquakes. *Geophysical Research Letters*, *41*(23), 8315–8324. Retrieved 2022-08-09, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2014GL061916 doi: 10.1002/2014GL061916

Poli, P., & Prieto, G. A. (2016, December). Global rupture parameters for deep and intermediate-depth earthquakes. *Journal of Geophysical Research: Solid Earth*, *121*(12), 8871–8887. Retrieved 2022-07-26, from https://onlinelibrary.wiley.com/doi/10.1002/2016JB013521 doi: 10.1002/2016JB013521

Prieto, G. A. (2022, May). The *Multitaper* Spectrum Analysis Package in Python. *Seismological Research Letters*, *93*(3), 1922–1929. Retrieved 2022-08-22, from https://pubs.geoscienceworld.org/srl/article/93/3/1922/612834/The-Multitaper-Spectrum-Analysis-Package-in-Python doi: 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. Retrieved 2022-10-11, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058109 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013GL058109) doi: 10.1002/2013GL058109

Prieto, G. A., Froment, B., Yu, C., Poli, P., & Abercrombie, R. (2017, March). Earthquake rupture below the brittle-ductile transition in continental lithospheric mantle. *Science Advances*, *3*(3), e1602642. Retrieved 2023-06-21, from https://www.science.org/doi/full/10.1126/sciadv.1602642 (Publisher: American Association for the Advancement of Science) doi: 10.1126/sciadv.1602642

Prieto, G. A., Parker, R. L., & Vernon III, F. L. (2009, August). A Fortran 90 library for multitaper spectrum analysis. *Computers & Geosciences*, *35*(8), 1701–1710. Retrieved 2023-05-01, from https://www.sciencedirect.com/science/article/pii/S0098300409000077 doi: 10.1016/j.cageo.2008.06.007

Radulian, M., & Popa, M. (1996, August). Scaling of source parameters for Vrancea (Romania) intermediate depth earthquakes. *Tectonophysics*, *261*(1), 67–81. Retrieved 2023-05-18, from https://www.sciencedirect.com/science/article/pii/0040195196000571 doi: 10.1016/0040-1951(96)00057-1

Rice, J. R. (2006). Heating and weakening of faults during earthquake slip. *Journal of Geophysical Research: Solid Earth*, *111*(B5). Retrieved 2023-05-29, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2005JB004006 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005JB004006) doi: 10.1029/2005JB004006

Schmandt, B., Jacobsen, S. D., Becker, T. W., Liu, Z., & Dueker, K. G. (2014, June). Dehydration melting at the top of the lower mantle. *Science*, *344* (6189), 1265–1268. Retrieved 2023-02-23, from https://www.science.org/doi/full/10.1126/science.1253358 (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1253358

Shi, Q. (2023, April). *TeleseismicDenoiser: DenoTe.* Zenodo. Retrieved 2023-06-01, from https://zenodo.org/record/7807794 doi: 10.5281/zenodo.7807794

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. Retrieved 2023-05-31, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2020GL089827 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL089827) doi: 10.1029/2020GL089827

Singh, S. K., Pacheco, J. F., Bansal, B. K., Pérez-Campos, X., Dattatrayam, R. S., & Suresh, G. (2004, August). 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. Retrieved 2023-06-01, from https://doi.org/10.1785/012003212 doi: 10.1785/012003212

Sobolev, A. V., Asafov, E. V., Gurenko, A. A., Arndt, N. T., Batanova, V. G., Portnyagin, M. V., … Byerly, G. R. (2019, July). Deep hydrous mantle reservoir provides evidence for crustal recycling before 3.3 billion years ago. *Nature*, *571* (7766), 555–559. Retrieved 2023-02-23, from https://www.nature.com/articles/s41586-019-1399-5 (Number: 7766 Publisher: Nature Publishing Group) doi: 10.1038/s41586-019-1399-5

Steblov, G. M., Ekström, G., Kogan, M. G., Freymueller, J. T., Titkov, N. N., Vasilenko, N. F., … Kondratyev, M. N. (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. Retrieved 2023-01-31, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2014GL060003 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014GL060003) doi: 10.1002/2014GL060003

Stockwell, R., Mansinha, L., & Lowe, R. (1996, April). Localization of the complex spectrum: the S transform. *IEEE Transactions on Signal Processing*, *44* (4), 998–1001. (Conference Name: IEEE Transactions on Signal Processing) doi: 10.1109/78.492555

Thomson, D. (1982, September). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, *70* (9), 1055–1096. (Conference Name: Proceedings of the IEEE) doi: 10.1109/PROC.1982.12433

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: Solid Earth*, *108* (B2). Retrieved 2023-06-01, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2002JB001948 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2002JB001948) doi: 10.1029/2002JB001948

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). Retrieved 2023-05-31, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2005GL023813 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005GL023813) doi: 10.1029/2005GL023813

Tschauner, O., Huang, S., Greenberg, E., Prakapenka, V. B., Ma, C., Rossman, G. R., … Tait, K. (2018, March). Ice-VII inclusions in diamonds: Evidence for aqueous fluid in Earth's deep mantle. *Science*, *359* (6380), 1136–1139. Retrieved 2023-02-23, from https://www.science.org/doi/full/10.1126/science.aao3030 (Publisher: American Association for the Advancement of Science) doi: 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. Retrieved 2023-05-01, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2022GL098402 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022GL098402) doi: 10.1029/2022GL098402

Vallée, M. (2013, October). Source time function properties indicate a strain drop independent of earthquake depth and magnitude. *Nature Communications*, *4*(1), 2606. Retrieved 2023-03-15, from https://www.nature.com/articles/ncomms3606 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/ncomms3606

Venkataraman, A., & Kanamori, H. (2004). Observational constraints on the fracture energy of subduction zone earthquakes. *Journal of Geophysical Research: Solid Earth*, *109*(B5). Retrieved 2023-05-01, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2003JB002549 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2003JB002549) doi: 10.1029/2003JB002549

Viesca, R. C., & Garagash, D. I. (2015, November). Ubiquitous weakening of faults due to thermal pressurization. *Nature Geoscience*, *8*(11), 875–879. Retrieved 2022-12-04, from https://www.nature.com/articles/ngeo2554 (Number: 11 Publisher: Nature Publishing Group) doi: 10.1038/ngeo2554

Warren, L. M., & Shearer, P. M. (2006, January). Systematic determination of earthquake rupture directivity and fault planes from analysis of long-period P-wave spectra. *Geophysical Journal International*, *164*(1), 46–62. Retrieved 2023-06-22, from https://doi.org/10.1111/j.1365-246X.2005.02769.x doi: 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. Retrieved 2023-05-31, from https://onlinelibrary.wiley.com/doi/abs/10.1002/grl.50977 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/grl.50977) doi: 10.1002/grl.50977

Wiens, D. A. (2001, December). 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. Retrieved 2023-06-01, from https://www.sciencedirect.com/science/article/pii/S0031920101002254 doi: 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: Solid Earth*, *108*(B4). Retrieved 2023-02-23, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2002JB001918 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2002JB001918) doi: 10.1029/2002JB001918

Ye, L., Lay, T., & Kanamori, H. (2020, November). Anomalously low aftershock productivity of the 2019 MW 8.0 energetic intermediate-depth faulting beneath Peru. *Earth and Planetary Science Letters*, *549*, 116528. Retrieved 2023-05-18, from https://www.sciencedirect.com/science/article/pii/S0012821X20304726 doi: 10.1016/j.epsl.2020.116528

Ye, L., Lay, T., Kanamori, H., Zhan, Z., & Duputel, Z. (2016, June). Diverse rupture processes in the 2015 Peru deep earthquake doublet. *Science Advances*, *2*(6), e1600581. Retrieved 2023-01-31, from https://www.science.org/doi/full/10.1126/sciadv.1600581 (Publisher: American Association for the Advancement of Science) doi: 10.1126/sciadv.1600581

Yin, J., Denolle, M. A., & He, B. (2022, December). A multitask encoder–decoder to separate earthquake and ambient noise signal in seismograms. *Geophys-*

ical Journal International, *231*(3), 1806–1822. Retrieved 2022-09-25, from https://doi.org/10.1093/gji/ggac290 doi: 10.1093/gji/ggac290

Yin, J., Li, Z., & Denolle, M. A. (2021, March). Source Time Function Clustering Reveals Patterns in Earthquake Dynamics. *Seismological Research Letters*, *92*(4), 2343–2353. Retrieved 2023-05-31, from https://doi.org/10.1785/0220200403 doi: 10.1785/0220200403

Zhan, Z. (2020). Mechanisms and Implications of Deep Earthquakes. *Annual Review of Earth and Planetary Sciences*, *48*(1), 147–174. Retrieved 2022-10-04, from https://doi.org/10.1146/annurev-earth-053018-060314 (_eprint: https://doi.org/10.1146/annurev-earth-053018-060314) doi: 10.1146/annurev-earth-053018-060314

Zhan, Z., Helmberger, D. V., Kanamori, H., & Shearer, P. M. (2014, July). Supershear rupture in a Mw 6.7 aftershock of the 2013 Sea of Okhotsk earthquake. *Science*, *345*(6193), 204–207. Retrieved 2023-06-22, from https://www.science.org/doi/full/10.1126/science.1252717 (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1252717

Zhan, Z., Kanamori, H., Tsai, V. C., Helmberger, D. V., & Wei, S. (2014, January). Rupture complexity of the 1994 Bolivia and 2013 Sea of Okhotsk deep earthquakes. *Earth and Planetary Science Letters*, *385*, 89–96. Retrieved 2023-06-22, from https://www.sciencedirect.com/science/article/pii/S0012821X13005979 doi: 10.1016/j.epsl.2013.10.028

Zhang, R. (2019, June). *Making Convolutional Networks Shift-Invariant Again.* arXiv. Retrieved 2023-02-15, from http://arxiv.org/abs/1904.11486 (arXiv:1904.11486 [cs]) doi: 10.48550/arXiv.1904.11486

Zhu, W., Mousavi, S. M., & Beroza, G. C. (2019, November). Seismic Signal Denoising and Decomposition Using Deep Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9476–9488. Retrieved 2022-08-11, from https://ieeexplore.ieee.org/document/8802278/ doi: 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. Retrieved 2022-06-15, from https://linkinghub.elsevier.com/retrieve/pii/S0065268720300030 doi: 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. Retrieved 2023-06-01, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2013JB010116 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013JB010116) doi: 10.1002/2013JB010116