REPORT
doi:10.26443/seismica.v5i1.1616

Harmonization of KiK-net and K-NET Ground-Motion Parameters and Metadata: A Flatfile of 950,000 Three-Component Records Spanning 29 Years

Karina Loviknes * [1,2,3], **Sebastian von Specht** [1], **Henning Lilienkamp** [1], **Annabel Händel** [1], **Chuanbin Zhu** [4], **Fabrice Cotton** [1,2]

[1]GFZ Helmholtz Centre for Geosciences, Potsdam, Germany, [2]University of Potsdam, Potsdam, Germany, [3]Njord Centre, Department of Physics and Geosciences, University of Oslo, Oslo, Norway, [4]School of Engineering, Physics and Mathematics, Northumbria University, Newcastle, UK

Author contributions: *Conceptualization*: Karina Loviknes, Sebastian von Specht, Fabrice Cotton. *Methodology*: Karina Loviknes, Sebastian von Specht, Henning Lilienkamp. *Writing - Original draft*: Karina Loviknes. *Editing*: Sebastian von Specht, Annabel Händel, Chuanbin Zhu, and Fabrice Cotton.

Abstract We present a new flatfile of automatically processed ground motions recorded by both the Kyoshin (K-NET) and Kiban–Kyoshin (KiK-net) networks in Japan. The flatfile comprises over 950,000 three-component ground motions recorded over 29 years (between May 1996 and December 2025), including the 2024 and 2025 $M_W > 7$ Noto, Hyuga-nada, and Aomori earthquakes. The ground motions were recorded by 1,047 K-NET surface stations and 702 KiK-net surface and borehole stations, with events ranging from JMA magnitude 2.5 to 9 and distances up to 1,200 km. The records are systematically processed, with detailed metadata describing both event parameters and recording characteristics. The flatfile includes key ground-motion parameters such as peak amplitudes (PGA, PGV, PGD), pseudo-spectral and Fourier amplitudes, and quality indicators like signal-to-noise ratios and usable frequency ranges. This flatfile is the first to integrate K-NET and KiK-net data at this scale, offering a harmonized and comprehensive resource for earthquake engineering research. Hosted in an open repository, it supports diverse applications, including ground-motion modeling, machine learning, and systematic analyses of ground-motion characteristics. The flatfile is available at https://doi.org/10.5880/GFZ.LKUT.2026.003.

Production Editor:
Andrea Llenos
Handling Editor:
Pablo Heresi
Copy & Layout Editor:
Hannah F. Mark

Signed reviewer(s):
Debi Kilb

Received:
February 28, 2025
Accepted:
May 22, 2026
Published:
June 19, 2026

1 Introduction

For systematic analysis of strong-ground motion data and development of ground motion models (GMMs), large and uniformly processed ground-motion datasets, covering a wide range of magnitudes and distances, are essential. A ground motion flatfile is a parametric table containing processed earthquake waveform parameters, metadata, and intensity measures. Widely used flatfiles include NGA-West 1 (Next Generation Attenuation Relationships for Western US, Chiou and Youngs, 2008) and 2 (Ancheta et al., 2014) for global shallow crustal earthquakes in active tectonic regions; RESORCE (Reference Database for Seismic Ground-Motion in Europe, Akkar et al., 2014) and ESM (European Engineering Strong-Motion, Lanzano et al., 2019; Luzi et al., 2020) for Europe; and the works of Dawood et al. (2016) and Bahrampouri et al. (2020) for Japan. More recently, country-specific flatfiles have been developed for e.g. Chile (Bastías and Montalva, 2016; Castro et al., 2022; Bastías et al., 2026), New Zealand (Manea et al., 2023), Türkiye (Sandıkkaya et al., 2024), and France (Traversa et al., 2020; Buscetti et al., 2024).

Japan hosts one of the world's most extensive, continuously operating seismic networks, with the Kyoshin (K-NET) and Kiban–Kyoshin (KiK-net) networks operated by the National Research Institute for Earth Science and Disaster Resilience (NIED) (2019). These networks provide high-quality ground-motion data from 1,047 surface stations (K-NET) and 702 surface and borehole stations (KiK-net) (Aoi et al., 2011; Okada et al., 2004), and have added incredible value for strong-motion research worldwide (e.g., Zhao et al., 2006; Kokusho and Sato, 2008; Bonilla et al., 2011; Dawood and Rodriguez-Marek, 2013; Kaklamanos et al., 2013; Wang et al., 2023). The large scale and high quality of the data have also allowed for training and testing of new methods that require large amounts of data, e.g. machine learning methods (e.g., Derras et al., 2020; Roten and Olsen, 2021; Esfahani et al., 2022; Lilienkamp et al., 2022). A few years ago, building upon the work of Dawood et al. (2016) (hereafter referred to as DAW16), Bahrampouri et al. (2020) (hereafter BEA20) developed an updated flatfile for KiK-net ground motions including a wider magnitude range, a new automatic processing scheme, and new ground motions recorded up to the end of 2017.

Since then, over 4,000 additional earthquakes have been recorded by KiK-net and K-NET, including several major events: the 2018 M_W 6.6 Hokkaido Eastern Iburi Earthquake (Ito et al., 2021), the 2022 M_W 7.4 Fukushima Earthquake (Kobayashi et al., 2023), the 2024 M_W 7.5 Noto and M_W 7.1 Hyuga-nada Earthquakes (Fujii and Satake, 2024; Itoh, 2025), and the 2025 M_W

*Corresponding author: karina.loviknes@mn.uio.no

7.6 Aomori earthquake (Yoshida et al., 2026). Here we provide a significantly expanded flatfile that includes records through December 2025 from both KiK-net and K-NET (DAW16 and BEA20 considered ground-motions from KiK-net only). The processing steps largely follow the methods of Boore (2005) and Ancheta et al. (2014), with some processing steps (e.g. selection of cutoff frequencies) differing from the previous datasets. The flatfile contains an extensive and more complete set of ground motion intensity measures, such as RotD50 pseudo-spectral amplitudes (PSA); Fourier amplitude spectra (FAS); peak ground acceleration (PGA), velocity (PGV), and displacement (PGD); Arias intensity; significant duration; and cumulative absolute velocity (CAV). In addition, we include several quality measures, including usable frequency ranges, signal-to-noise ratios, and an energy ratio parameter for detecting multiple event arrivals.

This flatfile is therefore well-suited for earthquake engineering research and systematic analysis of ground motions, particularly on K-NET, and large scale comparison between KiK-net and K-NET, which has not been possible with previous ground motion datasets (Dawood et al., 2016; Bahrampouri et al., 2020). Furthermore, because the waveform processing is fully automatic, the flatfile can be continuously updated with new ground motion records. In the following sections, the automatic processing steps are described and the resulting ground motion parameters are compared to other similar datasets.

2 Earthquake databases in Japan: NIED, JMA, and F-net

Japan is one of the most seismically active regions globally and has deployed strong-motion stations since the 1950s (Aoi et al., 2011). Following the 1995 Great Hanshin Earthquake, a major initiative was undertaken to establish a dense and uniformly distributed network of seismic stations capable of recording a broad range of amplitudes and frequencies. This led to the development of the broadband F-net, the high-sensitivity Hi-net, and the strong-motion networks K-NET and KiK-net (Aoi et al., 2011; Okada et al., 2004).

The K-NET network consists of 1,047 surface stations, uniformly spaced at approximately 20 km intervals, primarily in urban areas. In contrast, the KiK-net network comprises 702 co-located surface and borehole stations, integrated with Hi-net. In terms of site conditions, K-NET stations are generally located on softer soils with deeper sediments, as indicated by the distributions of the average shear-wave velocity in the top 10 m (V_{S10}) and fundamental frequency (f_0), shown in Figure 1a and b. Although V_{S30} is a more commonly used site parameter, it is not readily available for K-NET sites; therefore, V_{S10} is used here for comparison. The site parameters are retrieved from the site dataset of Zhu et al. (2021) which consists of a large variety of site parameters for the KiK-net and K-NET stations. The spatial distribution and event density for both networks are shown in Figure 1c and d.

The borehole depth and the V_S at the KiK-net borehole sensor are shown in Figure 1e and f. Although the KiK-net boreholes are at least 100 m deep and were designed to reach engineering bedrock, Figure 1f shows that this is not always achieved with $V_{S,\mathrm{borehole}}$ as low as 300 m/s for a few sites.

In addition to strong-motion data from KiK-net and K-NET, we incorporate earthquake information from the Japan Meteorological Agency (JMA) catalog and moment tensor solutions from the F-net catalog. Because KiK-net and K-NET metadata provide origin times only to the nearest minute, precise origin times from JMA and F-net are used. While both catalogs contain identical hypocenter information, the JMA catalog provides higher completeness, whereas the F-net catalog includes moment tensor solutions (Fukuyama, 1998). To ensure accurate event matching, we associate earthquakes recorded by KiK-net and K-NET with both catalogs using the origin times (in UTC), JMA magnitude (M_{JMA}), epicenter location, and distance in terms of surface rupture length surrounding the epicenters. The rupture length is estimated from the magnitude-scaling relation of Blaser et al. (2010). Although the estimated rupture length has uncertainty related to magnitude errors and misclassifications of focal mechanism, applying a fixed distance-window criterion would fail to account for the wide magnitude range in this dataset.

Events are considered matches when their differences in origin time are less than one minute, magnitude differences are within 0.3 units, and distances between K-NET/KiK-net and JMA epicenters are within three times the estimated rupture length. Due to uncertainties in depth estimation and the possibility that some events may have been assigned a standard depth (e.g., 320 events have depth =0 m), we do not include depth as a matching criterion.

The final JMA Unified Hypocenter Catalog is available until December 2023, while the F-net catalog includes events from January 1997 to December 2025. Between May 11, 1996, and December 31, 2025, a total of 20,742 earthquakes were recorded by KiK-net and K-NET. Of these:

- 17,941 (86.5%) match events in the JMA catalog (up to December 31, 2023)
- 12,414 (59.8%) match events in the F-net catalog (up to December 31, 2025)

For processing, we retain only events matched in either (or both) the JMA or F-net catalogs, excluding unmatched records. The final flatfile includes all available magnitude scales ($\mathrm{M_{JMA}}$, $\mathrm{M_W}$, $\mathrm{M_L}$, etc.) and depth values from both catalogs. The distributions of magnitude, depth, and event year for the included earthquakes of each catalog are shown in Figure S1 in Supplementary Material.

2.1 Focal mechanisms

In this study, we use moment tensor solutions from F-net (Fukuyama, 1998) and classify earthquakes as reverse (R), strike-slip (S), or normal (N) based on the dip angles of their respective axes relative to the horizontal, following the criteria of Boore and Atkinson (2007),

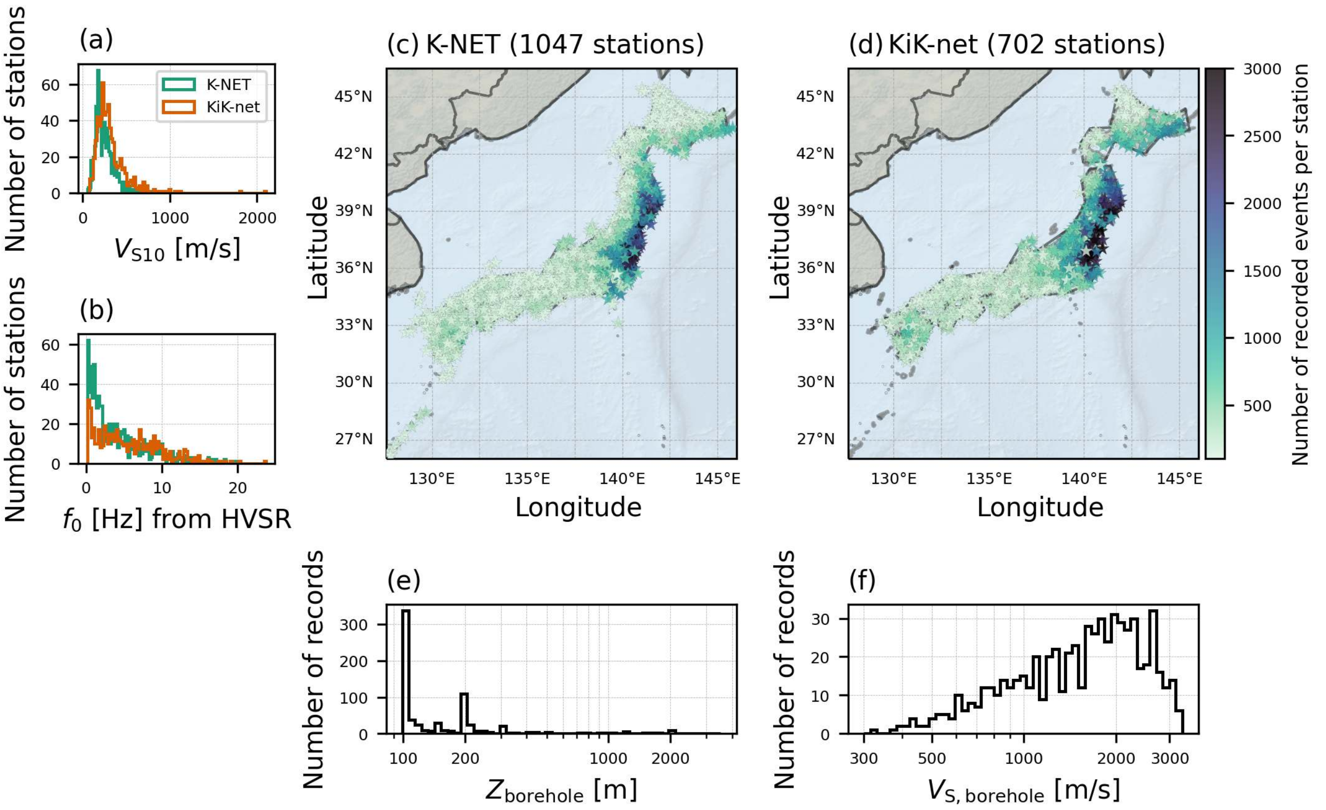

Figure 1 Comparison of K-NET (green) and KiK-net (orange) site characteristics. (a) Distribution of V_{S10} and (b) fundamental frequency (f_0), both obtained from the site dataset of Zhu et al. (2021). (c) Spatial distribution of K-NET stations and (d) KiK-net stations, with colors indicating the number of recorded events per station. (e) The distribution of depth to borehole, note that by design most boreholes are 100 or 200 m deep (Okada et al., 2004). (f) The distribution of V_S at the borehole sensor for the KiK-net borehole stations.

as summarized in Table 1. An event remains undefined (none) when not meeting any of the criteria set in Tab. 1 or when the moment tensor solution is not available (no F-net match).

Focal mechanism	P-axis plunge	T-axis plunge
Reverse (R)	$\leq 40°$	$> 40°$
Strike-slip (S)	$\leq 40°$	$\leq 40°$
Normal (N)	$> 40°$	$\leq 40°$

Table 1 Fault-type definition based on the criteria of Boore and Atkinson (2007).

The number of events per focal mechanism and the spatial distribution of the classified events are shown in Figure 2. The majority of recorded events are classified as reverse (R) or undefined (none, mainly due to no match with F-net catalog), which is consistent with Dawood et al. (2016) and Bahrampouri et al. (2020).

2.2 Distance Metrics and Rupture Plane Estimation

Epicentral (R_{epi}) and hypocentral (R_{hypo}) distances, representing the distances between a site and the earthquake epicenter and hypocenter, respectively, are computed for every record in the dataset. The Joyner-Boore (R_{JB}) and rupture plane (R_{rup}) distances are calculated only for events matched with the F-net catalog. These distances measure the shortest distance from a site to the rupture plane, with R_{rup} being the minimum distance to the rupture surface and R_{JB} representing the shortest distance to the surface projection of the rupture plane.

The rupture planes used to compute R_{JB} and R_{rup} are either directly from NIED (National Research Institute for Earth Science and Disaster Resilience, 2025) rupture models or estimated from the hypocenter location, $\mathrm{M_W}$, fault strike, dip, and rake from the F-net catalog (Fukuyama, 1998). The NIED rupture models are available for 12 significant earthquakes which are listed in Table S1 in Supplementary Material. For all other events that are matched with the F-net catalog, the rupture dimensions are estimated using the fault scaling relations of Blaser et al. (2010). The minimum distance is then determined using an iteratively refined mesh. All estimated rupture planes account for Earth's curvature, and are verified to ensure that calculated rupture planes do not extend above the Earth's surface; if necessary, the plane is adjusted down-dip to position the centroid just below the seismogenic depth (Scherbaum et al., 2004).

Our calculated R_{JB} and R_{rup} distances for KiK-net records are compared against corresponding distances from the BEA20 and DAW16 datasets (Fig. 3). Because different studies employ varying methods to estimate

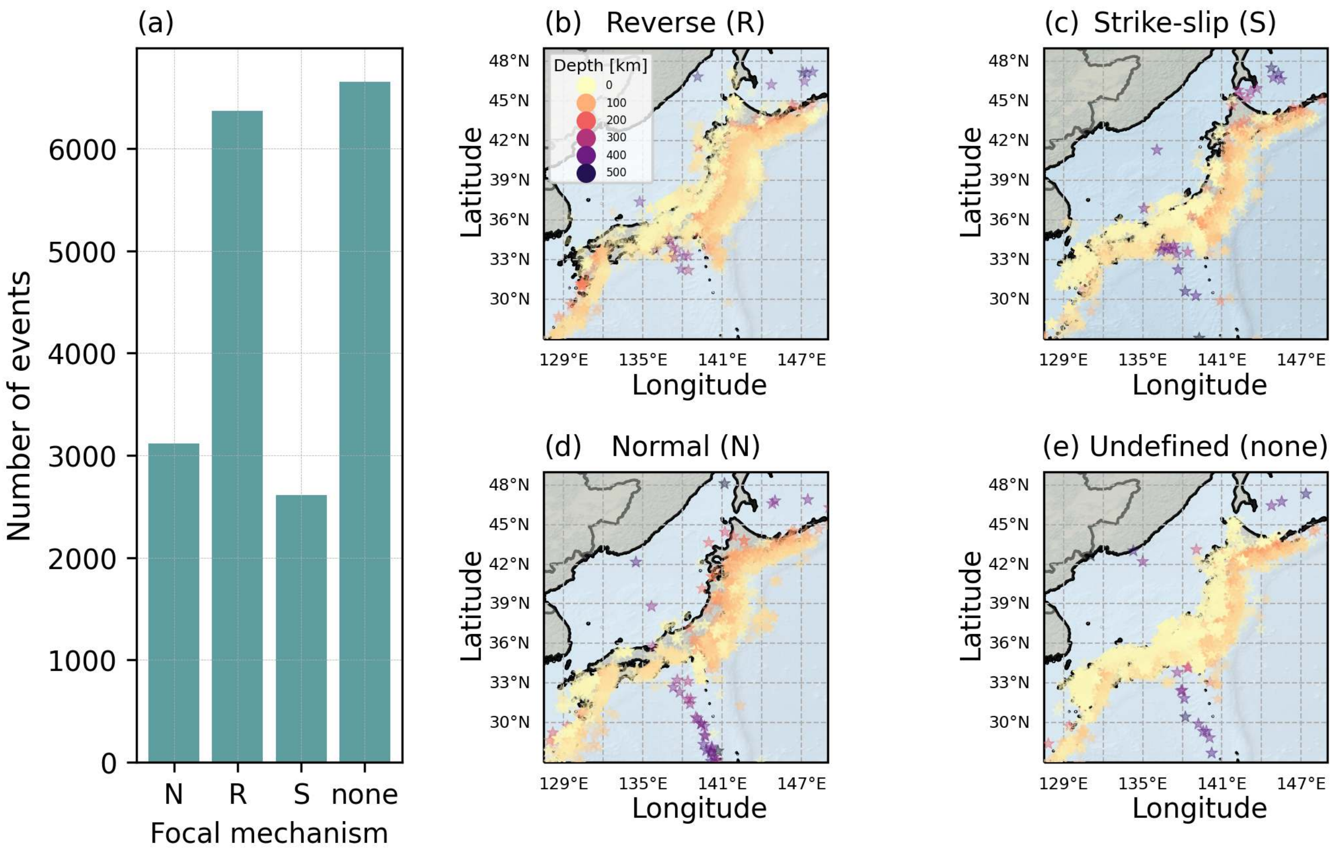

Figure 2 Focal mechanism classification of recorded events. (a) Number of events assigned to each focal mechanism. (b–e) Spatial distribution of events, colored by depth, and classified as (b) reverse (R), (c) strike-slip (S), (d) normal (N), and (e) undefined (none). The high number of undefined focal mechanisms is mainly related to events that are not included in the F-net catalog.

rupture planes and fault scaling, exact agreement between distance calculations is not expected. However, despite these methodological differences, the distributions of R_{JB} and R_{rup} are similar across datasets, with high Spearman correlation coefficients ($\rho > 0.96$). Nonetheless, Figure 3 show that notable differences exist, also between DAW16 and BEA20, and particularly for higher-magnitude events (darker colored points in Fig. 3). The primary source of these discrepancies is likely the method used to define rupture planes and select fault-scaling relations. Both DAW16 and BEA20 applied the approach of Chiou and Youngs (2008), where the Wells and Coppersmith (1994) scaling relation is used for all fault types. However, BEA20 further incorporated the relations of Thingbaijam et al. (2017), differentiating between shallow crustal and subduction earthquakes. In this study we employ the fault scaling of Blaser et al. (2010), which is based on a larger dataset than Wells and Coppersmith (1994) and also includes subduction events.

Another key distinction is the method used to determine the shortest distance to the rupture plane. This study refines a grid iteratively to locate the minimum distance, whereas DAW16 and BEA20 generated multiple (100–1000) rupture realizations and used the median distance. This difference contributes to deviations in the calculated distances, particularly for large rupture planes and high-magnitude events. Moreover, for larger earthquakes ($M_{\mathrm{JMA}} > 5$) and when available, rupture models from published sources are used. However, because multiple rupture models often exist for the same event, the different datasets do not necessarily use the same rupture models. This further explains why most of the events exhibiting large deviations are high-magnitude events, with the most significant differences observed for the M_{JMA} 9 Tohoku earthquake of March 11, 2011.

While these differences in R_{JB} and R_{rup} highlight the uncertainty associated with the calculations, it is beyond the scope of this study to draw conclusions about which method and results are most correct. The aim of this comparison is rather to demonstrate the overall consistency of results (as indicated by the high correlation coefficients) while acknowledging the inherent differences between approaches.

3 Automatic Processing of Acceleration Waveforms

The acceleration waveforms were processed automatically following the steps illustrated in Figure 4. The waveforms for each component were read, and the units were converted from cgs to SI, i.e., from gal to m/s^2. The data were then demeaned and detrended using a first-order polynomial to remove any linear trends.

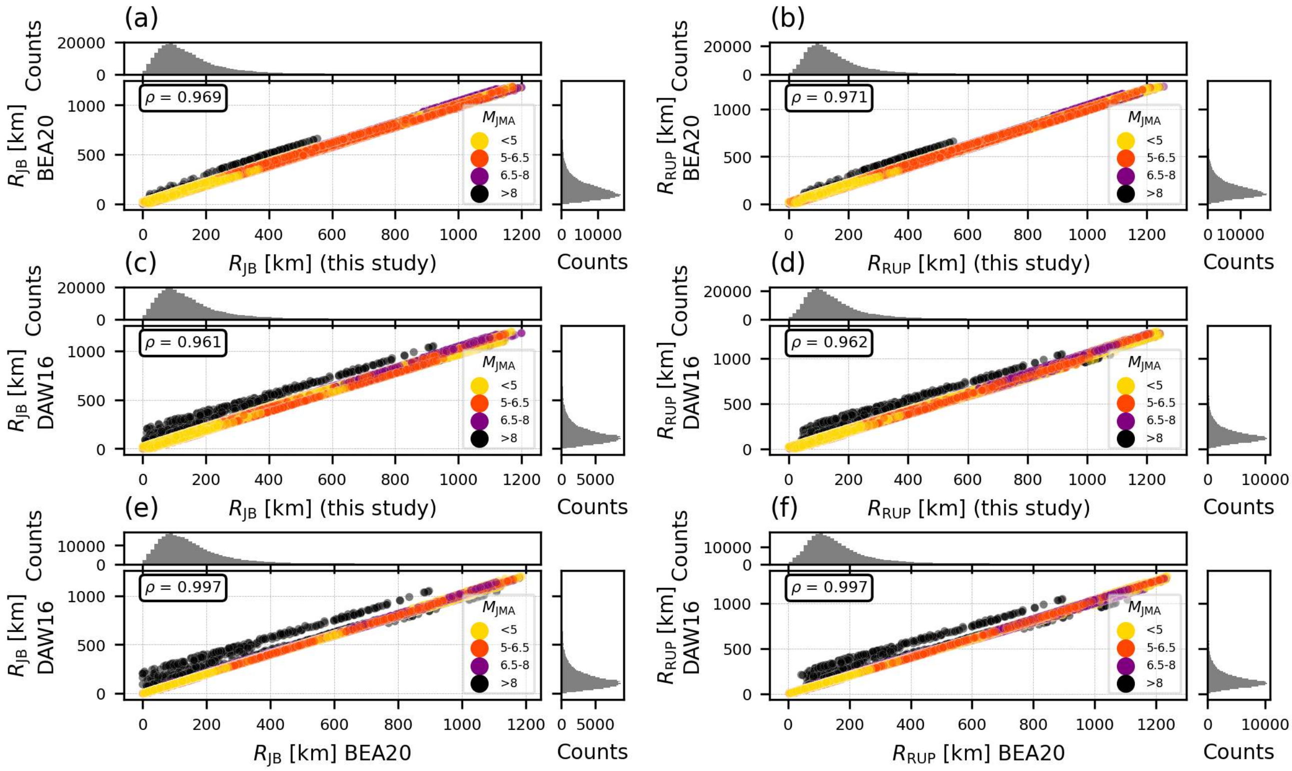

Figure 3 Distribution and comparison of R_{JB} (left) and R_{rup} (right) distances from KiK-net records. Colors indicate JMA magnitude. Top: Comparison between this study and BEA20. Middle: Comparison between this study and DAW16. Bottom: Comparison between BEA20 and DAW16. The correlation between distances in different studies is quantified using the Spearman correlation coefficient (ρ).

One of the primary challenges associated with the Kyoshin networks is that the records are trigger-based. Ideally, each record should begin at least 15 seconds before the P-wave arrival; however, this is not always the case. Some records start after the P-wave arrival or, in some cases, even after the S-wave arrival. Records that start after the S-wave arrival are discarded. To determine whether the P- and S-wave arrivals are within the recorded waveform, we use a travel-time table from JMA to calculate the theoretical P- and S-wave arrival times. The JMA travel-time table provides the P- and S-wave travel times as a function of focal depths and epicentral distance. Bilinear interpolation is then used to obtain any value within the table's range (distance: 0 km to 2000 km, depth: 0 km to 700 km). The travel time is then added to the event origin time to estimate the arrival times at the recording station. Since the origin times provided in the metadata are only accurate to the nearest minute, we use the origin times from the JMA catalog. If the theoretical P-wave arrival occurs significantly late after the start of the record (here, arbitrarily set to 80 s), the record is cut to 11 seconds before the theoretical P-arrival to retain only the relevant portion of the waveform.

3.1 Energy Ratio

Identifying waveforms containing multiple event arrivals is a major challenge in automatic processing. Such waveforms can cause high-frequency distortion and unreliable estimates of certain intensity measures, such as Arias intensity and CAV (Dawood et al., 2016; Bahrampouri et al., 2020).

To assess the potential for multiple event arrivals in the time series, we introduce a method that compares the computed energy within the theoretical P- and S-wave windows to the total energy of the entire record. In an undisturbed waveform, most of the energy (here set to 80%) should be within the theoretical P- and S-wave windows (e.g., Perron et al., 2018). If more than 20% of the total energy lies outside this window, we assume contamination by an unexpected signal, such as another event arrival. These records are still included in the flatfile but should be used with caution.

The theoretical S-wave window duration is calculated following Perron et al. (2018), using the corner frequency (f_c) derived from the Brune model (Brune, 1970):

$$f_c = 0.37 V_{S,\mathrm{crust}} \left(\frac{16 \Delta\sigma \cdot 10^5}{7 M_0} \right)^{1/3}, \tag{1}$$

where the average crustal shear-wave velocity ($V_{S,\mathrm{crust}}$) is set to 3500 m/s, the stress drop $\Delta\sigma$ = 1 MPa, and the seismic moment M_0 is computed from the moment magnitude M_W using the relation from Kanamori (1977):

$$M_0 = 10^{1.5 M_W + 9.1}. \tag{2}$$

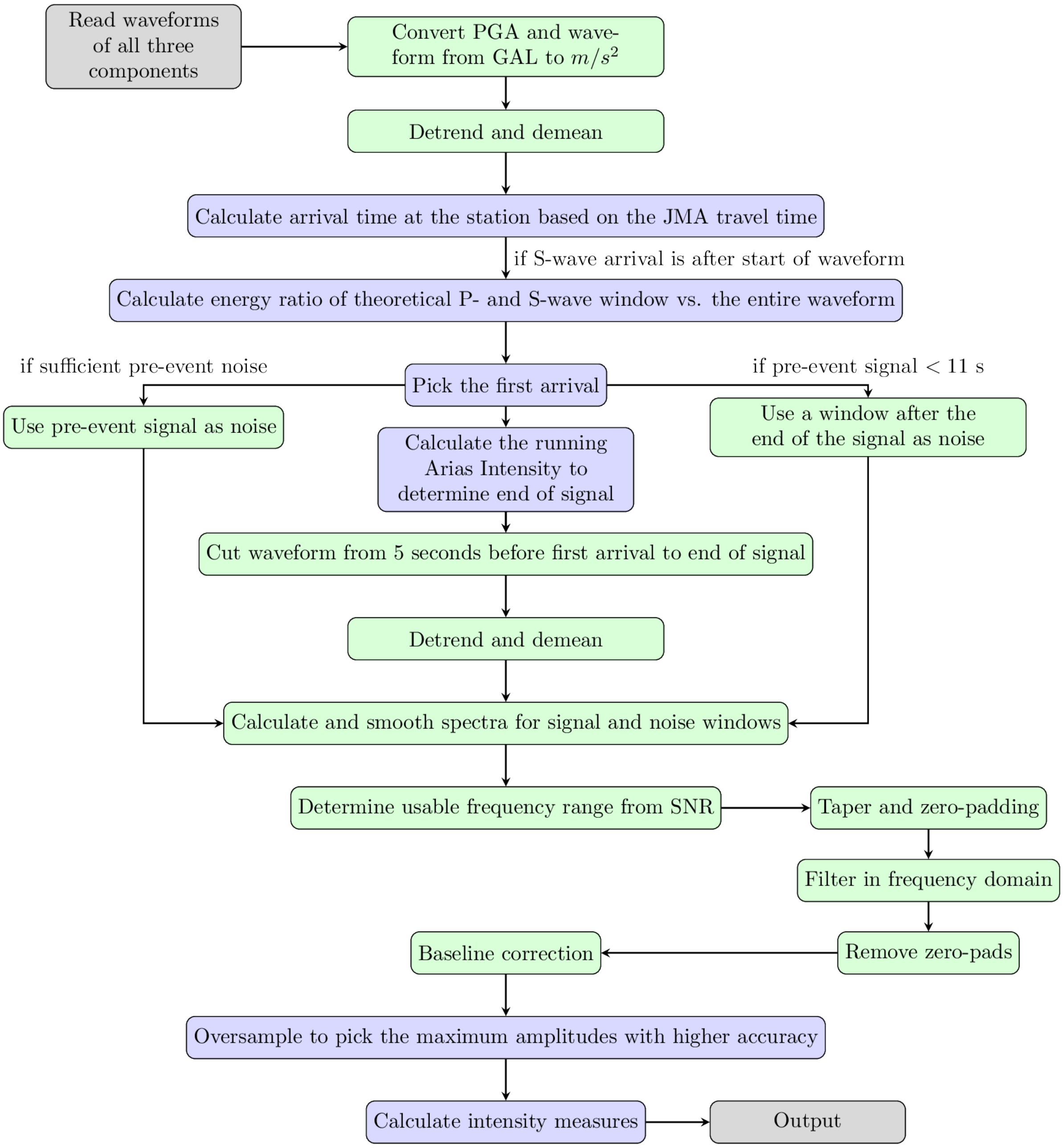

Figure 4 Flow chart of the processing steps. The grey boxes refers to input and output of the processing. The processing steps in the green boxes are applied to the waveforms, while the processing in the blue boxes are related to computations of values from the waveforms. Each processing step is described in more detail in the *Processing steps* section.

The theoretical S-wave window duration D_S is then given as:

$$D_S = \min\left[\max\left(D_{S,\min}, \frac{1}{f_c} + (T_S - T_P)\right), D_{S,\max}\right], \quad (3)$$

where T_P and T_S are the theoretical P- and S-wave arrival times obtained from the JMA travel-time table, and $D_{S,\min}$ and $D_{S,\max}$ are the minimum and maximum S-wave durations, respectively. $D_{S,\min}$ is here set to 5 s, to ensure sufficient spectral resolution at low frequencies (Perron et al., 2018), particularly for waveforms recorded at stations close to the hypocenter. $D_{S,\max}$ is set to 120 s as this was initially the minimum length of the raw waveforms from the Kyoshin networks (Aoi et al., 2011; Okada et al., 2004).

The energy of the theoretical P- and S-wave window (combined) and the energy of the entire waveform (of each component) are computed as the sum squared acceleration amplitudes. Finally, the ratio of signal-window energy to total energy is obtained for each component. If the mean energy ratio of the three components is less than 80%, the record is flagged as a potential multiple-event arrival, as demonstrated in Figure 5 for a waveform recorded on 2004-10-23 at station NIG028. It should be noted, however, that this method does not detect cases where multiple arrival events occur within the theoretical P- and S-wave window. For such cases, there is no simple and automatic way to detect multiple arrival events.

3.2 Determination of Signal and Noise Windows

The observed first arrival is automatically detected using the Short-Term Average over Long-Term Average (STA/LTA) method (Allen, 1978), following the approach of Akazawa (2004). The first arrival is identified as either P- or S-wave depending on whether the waveform starts before or after the theoretical P-wave arrival. If the first arrival is P-wave and at least 11 s of pre-event data are available, we use the pre-event window as noise. If the first arrival is S-wave or insufficient pre-event noise is available, the noise window is selected either as the last 20 seconds of the record, or (if the last 20 seconds contain excessive noise) a quieter noise window (at least 11 seconds long) is identified between the end of the main signal and the end of the record.

Identifying a suitable noise window unaffected by earthquake signals is crucial for accurately estimating the signal-to-noise ratio (SNR). Pre-event noise is preferred over post-event noise because KiK-net and K-NET records have limited durations (typically not exceeding 300 s), increasing the likelihood of coda wave contamination at the end of the signal. To evaluate the noise window quality, we compute the ratio of noise window energy to total waveform energy. If this noise energy ratio is low, we assume minimal earthquake signal contamination. A threshold of 1% is used, corresponding to the mean noise energy ratio for records with sufficient pre-event noise (at least 11 seconds before the P-wave arrival).

If a waveform lacks sufficient pre-event noise and the noise energy ratio of the post-signal window is too high (> 0.01), an alternative noise window is sought at the end of the signal. If no acceptable noise window is found, the record is flagged but still included in the flatfile and should be used with caution.

3.3 Significant Duration and End of Signal

The significant duration and end of the main signal are determined using the running Arias intensity (Arias, 1970; Bommer and Martínez-Pereira, 1999):

$$I_A(t) = \frac{\pi}{2g} \int_0^t a(\tau)^2 \, \mathrm{d}\tau, \tag{4}$$

where $a(\tau)$ is the acceleration at time τ and g is gravitational acceleration. The normalized Arias intensity, $\hat{I}_A(t) = I_A(t)/I_A(t_{\max})$, is computed, where $t_{\max}$ is the total waveform duration. Significant duration is defined as the time interval between the 5th and 75th percentile (D_{5-75}) and the 5th and 95th percentile (D_{5-95}) of $\hat{I}_A(t)$ (e.g., Bommer et al., 2009; Baker et al., 2021). The end of the main signal is taken as the maximum of the 97.5th percentile ($D_{5-97.5}$) and the theoretical S-wave window duration D_S, as demonstrated in Figure S2 in Supplementary Material for a waveform of the Chuetsu-oki earthquake (23^{rd} of October 2004) at station SIT011. The significant duration values (D_{5-75}, D_{5-95}) which are included in the flatfile are computed from the final processed waveform.

The waveform is cut to include only the signal, retaining 5 seconds before the observed P-wave arrival to the identified end of the signal. This ensures sufficient data for subsequent tapering. If the observed P-wave arrival occurs before the record start, the signal is cut to begin 5 seconds before the S-wave arrival. Records starting after the S-wave arrival are omitted. After cutting, the waveform is again detrended using a 1st order polynomial.

3.4 Finding Usable Frequency Range and Selecting Filter Corner Frequencies

To calculate the signal-to-noise ratio (SNR) and determine the usable frequency band, the spectra for both the signal and noise windows are computed and smoothed. Spectral smoothing is performed using LOESS (locally estimated scatterplot smoothing) regression (Hastie et al., 2009), where a fourth-order polynomial is fitted over $k = 30$ nearest neighbors around each target frequency.

The usable frequency band is defined as the frequency range where the majority of the SNR values exceed 3 (Boore and Bommer, 2005), ignoring windows with fewer than five samples (Fig. 6). The filter corner frequencies are selected as the minimum and maximum frequencies within this usable range, constrained by predefined limits based on the noise window duration D_{noise} to avoid unrealistic values (Tab. 2).

Corner frequency	Minimum	Maximum
Lower cutoff frequency $f_{c,0}$	$f_{c,0,\text{min}} = 1/D_{noise}$	$f_{c,0,\text{max}} = 0.7$ Hz
Higher cutoff frequency $f_{c,1}$	$f_{c,1,\text{min}} = 15$ Hz	$f_{c,1,\text{max}} = 30$ Hz

Table 2 Minimum and maximum lower ($f_{c,0}$) and upper ($f_{c,1}$) corner frequencies used for filtering.

The inner frequency range (0.7 Hz to 15 Hz) is selected to ensure a sufficiently wide frequency band for each record, while $f_{c,1,\max}$ is constrained by the 30 Hz anti-aliasing filter applied to the data prior to recording (Aoi et al., 2011). Each corner frequency is determined separately for the horizontal components, and the more conservative of the two is selected, ($f_{c,0} = \max[f_{c,0,\mathrm{EW}}, f_{c,0,\mathrm{NS}}]$ and $f_{c,1} = \min[f_{c,1,\mathrm{EW}}, f_{c,1,\mathrm{NS}}]$).

Although using the most conservative frequency band may filter out valuable information from high SNR components due to an orthogonal component with low SNR, we use the same cut-off frequency for all components for two main reasons. Firstly, the pseudo-spectral amplitudes are computed from a rotational combination of the horizontal components. Using a combination of the horizontal components for ground motion analysis is common practice in earthquake engineering (e.g., Ancheta et al., 2014), where flatfiles are widely used. Secondly, the selected cutoff frequencies are subject to large uncertainty due to the aforementioned issue of finding a long, high-quality noise window.

If the selected $f_{c,0} > 0.7$ Hz, the record is flagged for potential low-frequency noise, and if $f_{c,1} < 15$ Hz, it is flagged for potential high-frequency noise. The usable

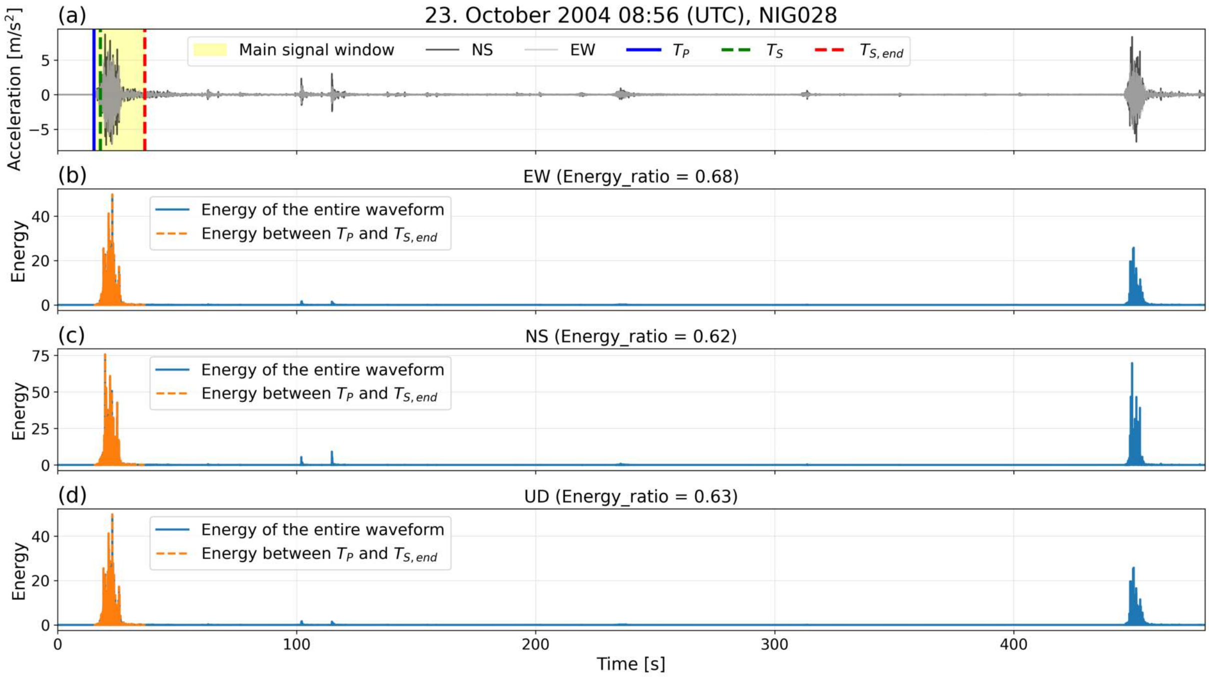

Figure 5 Demonstration of how the energy ratio detects multiple event arrivals in the waveform for the $M_W 6.8$ earthquake recorded on October 23, 2004, at station NIG028. (a) Detrended waveform for the NS (black) and EW (silver) components. (b–d) Energy ratio between the main signal window (P- and S-wave, dotted orange line) and the entire waveform (solid blue line) for the (b) EW, (c) NS, and (d) UD components.

frequency range ratio is computed as

$$q_{\text{range}} = \frac{f_{c,1} - f_{c,0}}{f_{c,1,\text{max}} - f_{c,0,\text{min}}}. \quad (5)$$

Records flagged for low or high-frequency noise, or with $q_{\text{range}} < 0.6$, should be used with caution. Additionally, the mean SNR are provided for each horizontal component. However, as discussed previously, SNR calculations depend on the selected noise window and should only be considered reliable for records with acceptable noise windows (i.e., with noise energy ratios < 0.01). Note that while we provide guidelines for identifying low-quality records, the flat file includes all records regardless of q_{range} and noise energy ratios. Therefore, data selection is left to the user.

3.5 Filtering and Baseline Correction

Prior to filtering, waveforms are tapered to reduce spectral leakage using a 2.5% cosine taper (Havskov and Ottemöller, 2010) and padded with zeros to prevent contamination from finite-duration signal effects and the circular properties of the Fast Fourier Transform (FFT) (Boore, 2005).

Filtering is performed using an acausal fourth-order Butterworth filter (implemented with the SCIPY Python package). After filtering, the zero-padding is removed, and a baseline correction is applied to eliminate non-physical long-term trends, ensuring compatibility between the processed acceleration, velocity, and displacement time series (Boore et al., 2012). The baseline correction follows the approach of Boore et al. (2012) and Ancheta et al. (2014).

First, the acceleration time series is double-integrated to displacement. A sixth-order polynomial, with constraints on the zeroth and first-order coefficients to prevent shifts in displacement and velocity, is then fitted to the displacement trace. Finally, the second derivative of this polynomial is subtracted from the acceleration trace, and the corrected acceleration trace is integrated to obtain the corrected velocity and displacement time series.

3.6 Calculation of Intensity Measures

In the final step, before calculating intensity measures, the baseline-corrected acceleration and velocity time series are oversampled to 400 Hz. Oversampling is performed using the Whittaker–Shannon interpolation formula (Shannon, 1949), also known as sinc interpolation, where the original time series is convolved with the sinc function at time t and time interval Δt. The purpose of oversampling is to estimate the largest potential extrema within the dataset (Nigam and Jennings, 1969; Boore and Goulet, 2014). While linear interpolation of the time-series could lead to underestimation of the absolute peaks and introduce high frequency noise above the Nyquist frequency (f_{Nyquist}), the sinc interpolation used here avoids these issues by zero-padding the Fourier spectra beyond f_{Nyquist} before transforming back to time domain (Boore and Goulet, 2014). To evaluate the impact of oversampling we compute the peak amplitudes before and after oversampling. The comparison, shown in Figure S3 in Supplementary Material, shows that the values are highly similar, with Spearman's correlation coefficients $\rho > 0.99$, demonstrating that oversampling does not introduce artifacts in the

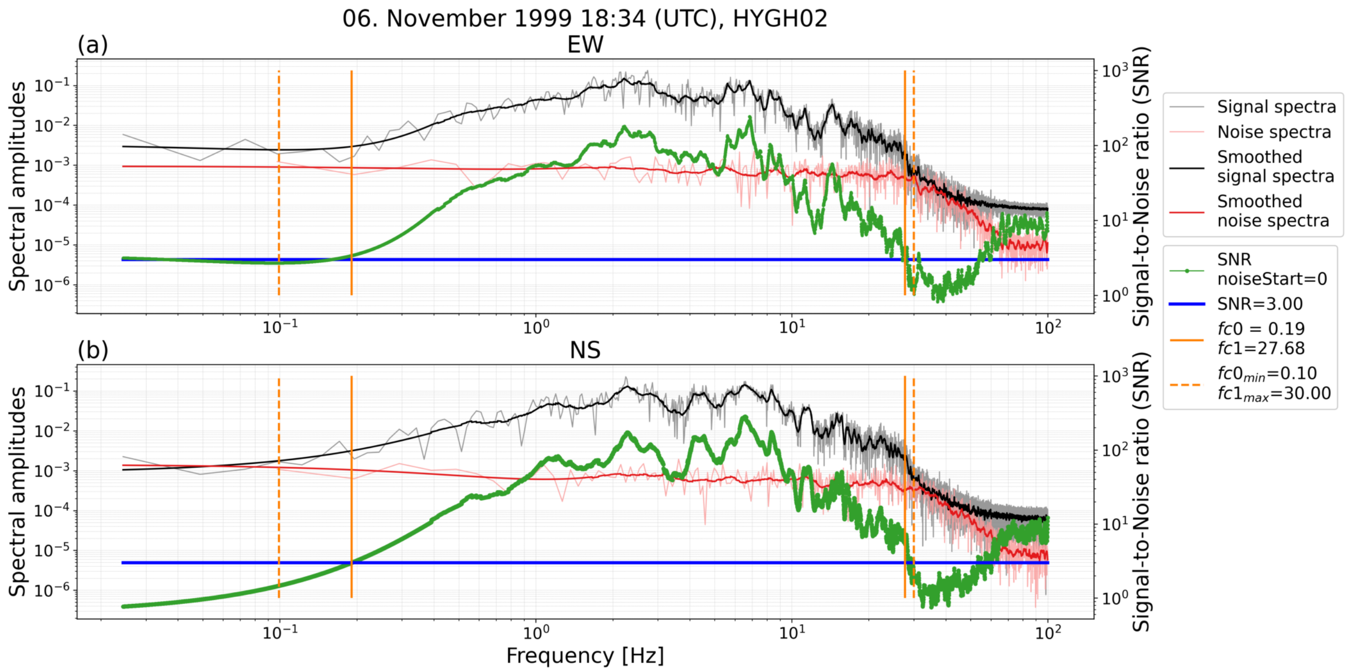

Figure 6 Selection of the usable frequency range and corner frequencies for filtering, demonstrated for the EW (top) and NS (bottom) components of the $M_W 4.8$ earthquake recorded on November 11, 1999, at station HYGH02. The signal and noise are smoothed using LOESS (locally estimated scatterplot smoothing) regression.

peak amplitude computation.

The peak amplitudes for acceleration (PGA), velocity (PGV), and displacement (PGD) are then identified as the absolute maxima of the oversampled traces for each horizontal component. Additionally, pseudo-spectral acceleration (PSA), Fourier amplitude spectra (FAS), Arias intensity (I_A), significant duration (D_{5-75}, D_{5-95}), and cumulative absolute velocity (CAV) are computed from the processed waveforms.

PSA, which represents the peak response of a single-degree-of-freedom (SDOF) harmonic oscillator subjected to a ground motion record, is calculated using the Nigam and Jennings (1968, 1969) algorithm. PSA values are computed for RotD0, RotD50, and RotD100, which are determined by rotating the two horizontal components in 1° increments through all angles (0° to 180°). At each angle, spectral acceleration is computed from the rotated time series, and RotD0, RotD50 and, RotD100 are defined as the minimum, median, and maximum, respectively, over the rotation angle for each response period T (Boore et al., 2006; Boore, 2010).

The FAS for all three components are calculated using the discrete Fourier transform for 239 logarithmically spaced frequencies between 0.1 Hz and 24 Hz. Arias intensity (I_A) and CAV are computed from the acceleration time series ($a(\tau)$, e.g., Baker et al., 2021):

$$I_A = \frac{\pi}{2g} \int_0^{t_{\max}} a(\tau)^2 \, d\tau \quad (6)$$

and

$$CAV = \int_0^{t_{\max}} |a(\tau)| \, d\tau, \quad (7)$$

where g is gravitational acceleration.

4 Processed Dataset Overview

Of the 20,742 earthquakes recorded by KiK-net and K-NET, 18,738 (90.3%) were successfully processed, resulting in 950,796 ground-motion records from 1,749 stations. Among the unprocessed events, the majority (1,571) were rejected due to no match with either the JMA or F-net catalogs, and 582 events were discarded due to low magnitude ($M_{JMA} \leq 2.5$). Additionally, 17,566 ground-motion records were rejected during processing, primarily due to waveforms starting after the S-wave arrival (Fig. S4a in Supplementary Material). The year, magnitude, distance, and PGA distributions of the rejected records are shown in Figure S4b–e. Notably, a larger number of records were rejected from KiK-net compared to K-NET. Additionally, as also shown in Figure S4e, the PGA distribution of the rejected KiK-net records is centered around lower PGA values ($\sim$0.003 m/s^2) than K-NET ($\sim$0.02 m/s^2). This is likely due to KiK-net recording lower-amplitude ground motions more often than K-NET, as a result of different thresholds used to trigger recording by the two networks (Aoi et al., 2004): while KiK-net starts recording at 0.002 m/s^2 to 0.004 m/s^2 (0.2 gal to 0.4 gal), K-NET has a higher threshold of 0.02 m/s^2 (2 gal).

The results of the processing and an overview of the parameters included in the dataset are given in this section. Because K-NET and KiK-net use the same instruments and are installed and organized by the same agency (National Research Institute for Earth Science and Disaster Resilience (NIED), 2019), the same processing steps were applied to both networks. In addition, the recordings from the KiK-net borehole stations are also processed. To efficiently process the borehole records, the borehole stations are considered as a separate network from the KiK-net surface stations in the processing procedure. While the overall processing results are similar across the three networks (K-NET, KiK-net surface and KiK-net borehole), some differences exist and are discussed in the following subsections.

4.1 Waveform Start Time and Noise Window Selection

A key challenge in processing Kyoshin network records is that they are trigger-based. Ideally, each record should start at least 15 seconds before the P-wave arrival; however, this is not always the case. Some records start after the P-wave arrival or even after the S-wave arrival. In the latter case, the record is discarded. While most ground-motion records start at the P-wave arrival, approximately 25 % start from the S-wave arrival across both KiK-net (surface and borehole) and K-NET (Fig. 7a). Additionally, as described in the *Processing Steps* section, records where the theoretical P-wave arrival occurs more than 80 s after the start are truncated to 11 s before the theoretical P-arrival. This is the case for 0.3% of the records (2,645 records; 2,040 from KiK-net and 605 from K-NET).

Late triggering is particularly challenging for SNR calculations in relation to selecting a suitable noise window unaffected by earthquake signals. As described in the *Processing Steps* section, a pre-event noise window is preferred but must be at least 11 seconds long. If this is not the case, a suitable noise window is selected between the end of the main signal (determined using the Perron et al., 2018, method) and the end of the record. The quality of the noise window is assessed based on the energy ratio between the noise window and the entire waveform. A noise window is considered acceptable if this energy ratio is below 1% (< 0.01).

Figure 7b shows that for KiK-net (surface and borehole), nearly half (45%) of the records have sufficient pre-event noise, with slightly more records from the borehole stations than from the surface stations. In contrast, most K-NET records rely on the end of the signal as the noise window. Furthermore, 60% of K-NET records have a noise energy ratio below 0.01, compared to 90% of KiK-net records (Fig. 7c). This discrepancy may be attributed to the shorter average length of K-NET records. Figure 7d shows that while most KiK-net records are 120 s long, the majority of K-NET records are 60 s in length. Consequently, noise windows selected from the end of these shorter records may be contaminated by earthquake signals, such as coda waves. For the final processed waveforms, however, K-NET and KiK-net show similar length distributions (Fig. 7e).

4.2 Frequency content, energy ratio and cut-off frequencies

The three main parameters for evaluating the usability and quality of each ground-motion record in this flatfile are the usable frequency range ratio q_{range} (eq. 5), the noise energy ratio, and the energy ratio of the main signal (P and S-wave window) relative to the entire waveform. Ground-motion records are considered acceptable quality if they meet the following criteria:

- Usable frequency range ratio: $q_{\text{range}} > 0.6$
- Noise energy ratio: < 0.01
- Signal energy ratio: > 0.8

Ground motions with values outside these ranges are still included in the flatfile, but should be used with caution. Figure 8a shows that most records have high q_{range}, although especially K-NET has a significant number (10%) of records with little to no usable frequency range. This is likely related to the difficulty of finding a suitable noise window, which is particularly challenging for the many short K-NET records (Fig. 7c and d). For KiK-net, the surface and borehole recordings have similar distribution, with slightly more borehole records having q_{range} close to one. The signal energy ratio distributions, however, are similar for both K-NET and KiK-net (surface and borehole), with high values (> 0.8) for most records. This shows that for most records the main energy is within the expected P- and S-wave window and not distorted by multiple events or unexpected noise sources. Considering the quality criteria described above 59% of the records (556,831 records; 303,143 from KiK-net and 253,688 from K-NET) can be considered high quality.

Figures 8c and d show the distributions of the low- ($f_{c,0}$) and high-cutoff frequencies ($f_{c,1}$) used for waveform filtering. As described in the *Processing steps* section and Table 2, the corner frequencies are constrained ($f_{c,0,\text{min}} = \frac{1}{D_{noise}}$ Hz, $f_{c,0,\text{max}} = 0.7$ Hz, $f_{c,1,\text{min}} = 15$ Hz and $f_{c,1,\text{max}} = 30$ Hz). The distributions in Figure 8c and d show that for the majority of the records the corner frequencies are the same as the constraining values. This is particularly relevant for $f_{c,0}$, suggesting that many records require a higher low-frequency cutoff due to low-frequency noise, such as instrumental or weather-related noise (e.g. Aoi et al., 2011). While the upper cutoff frequency $f_{c,1}$ is constrained by the 30 Hz anti-aliasing filter applied at KiK-net and K-NET sites (Aoi et al., 2011), a substantial number of records have $f_{c,1}$ set to the minimum 15 Hz (27% for KiK-net and 38% for K-NET), indicating potential high-frequency noise contamination. These records should therefore be used with caution. Note that because the KiK-net borehole records are processed separately from the surface records, the records are filtered using different cutoff frequencies. For site response studies using the surface-to-borehole-ratio (SBr, e.g. Bonilla et al., 2011; Régnier et al., 2013), only the most conservative frequency range ($f_{\text{SBr,min}} = \max[f_{c,0,\text{surface}}, f_{c,0,\text{borehole}}]$ and $f_{\text{SBr,max}} = \min[f_{c,1,\text{surface}}, f_{c,1,\text{borehole}}]$) should therefore be considered.

The distribution of the different types of noise windows used to select lower, mid, and maximum range of $f_{c,0}$ and $f_{c,1}$ are shown in Figures 8e–j. For $f_{c,0}$, the noise window seems to have an impact on the selected cutoff frequency, with the majority of the records with $f_{c,0} = 0.7$ Hz relying on a noise window from the end of the waveform (Fig. 8g), while pre-event noise is used for the majority of the records with $f_{c,0} < 0.7$ Hz (Figs. 8e and f). For $f_{c,1}$, the noise window does not appear to

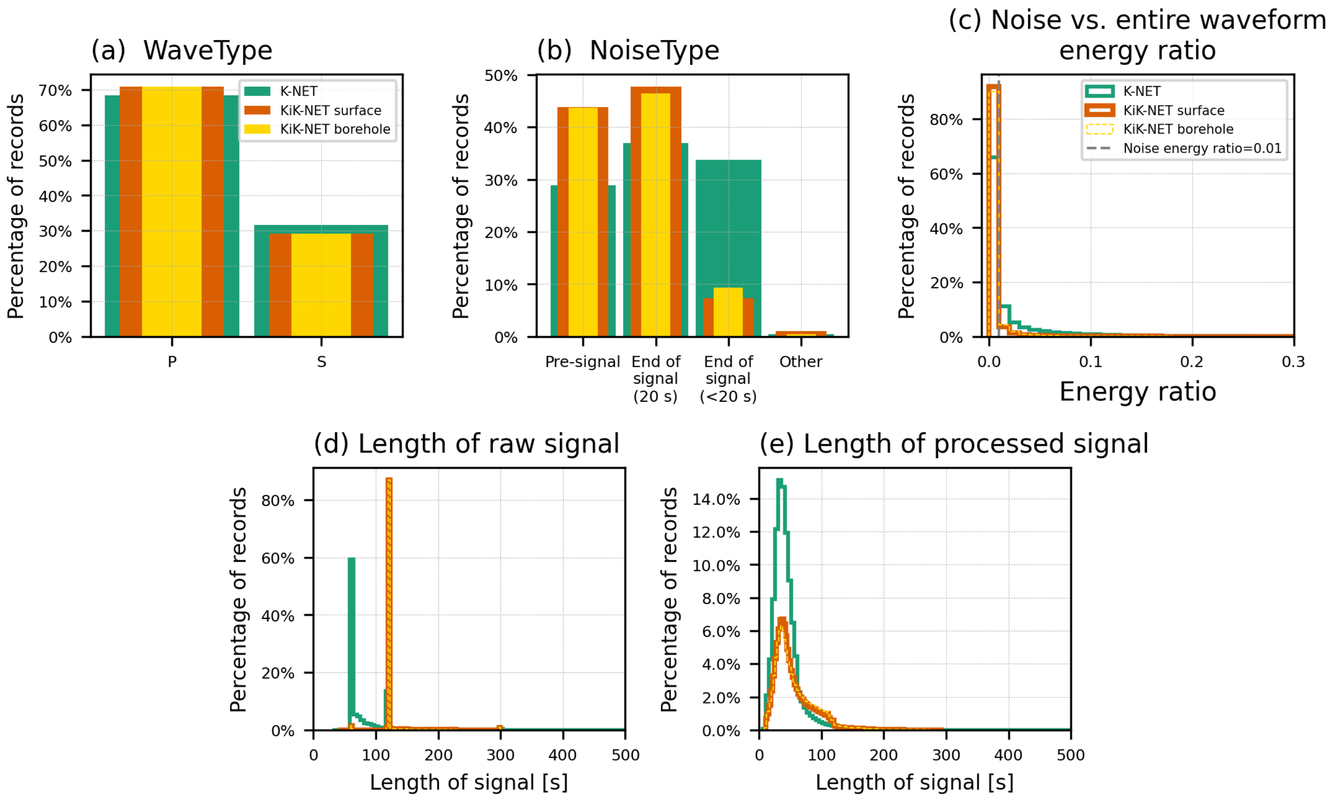

Figure 7 Distribution of waveform and noise characteristics for K-NET (green), KiK-net surface (orange), and KiK-net borehole (yellow). (a) Wave type (*P* for waveforms starting at the P-arrival and *S* for waveforms starting at the S-arrival). (b) Noise type. (c) Energy ratio between the noise window and the rest of the waveform. (d) Length of raw signal. (e) Length of final processed signal. The noise windows are separated into pre-event noise (prior to the P-arrival), and end of signal noise, either as exactly the last 20 seconds of the waveform, within the last 20 second but with length shorter than 20 s, or sometime between the end of the signal and end of the waveform (other).

have a strong impact on the selected cutoff frequency, with similar distribution for all ranges (Figs. 8h-j).

5 Comparison with Other Processed Flatfiles

As described in the *Introduction*, several other flatfiles of ground-motion records have been published in recent years, including datasets containing records from KiK-net and K-NET. Here, we compare the results of our processing with some of these studies. In addition to the KiK-net flatfiles from DAW16 and BEA20, we also compare our results with the NGA-West 2 dataset (Ancheta et al., 2014), which also includes K-NET records.

Similar to the comparison of distances in the *Distance Calculations* section, due to differences in computation methods and processing steps, we do not expect exact agreement between datasets (e.g., Boore et al., 2012). Here, we rather highlight the similarities and differences due to the different processing schemes, as previously discussed in similar studies (Paolucci et al., 2011; Boore et al., 2012; Lanzano et al., 2019). For comparisons with KiK-net, only surface records are considered.

Recently, a flatfile from NIED was published containing ground motion records recorded by KiK-net and K-NET up until the end of 2024 (Morikawa et al., 2024, https://www.j-shis.bosai.go.jp/en/labs/ground-motion-flatfile/, last accessed 18.02.2026). Because the flatfile was published only recently, we do not provide a comprehensive comparison here. However, the main differences between the two are that the flatfile from this study includes FAS, PGD, and distances based on rupture (R_{JB} and R_{rup}), while the NIED flatfile includes PGA, PGV, and PSA calculated for RotD25 and RotD75 (in addition to RotD0, RotD50 and RotD100) and classifications of earthquake types. Also, DAW16 and BEA20 include earthquake classifications, separating between interface subduction, intraslab subduction, outer subduction, and shallow crustal events. Although separating earthquake types is important for ground motion modeling, the method of classification is highly user-dependent. Therefore, we leave this step to the users of the flatfile.

The magnitude-distance-PGA distribution for K-NET and KiK-net (surface) in this study, NGA-West 2, DAW16, and BEA20 is shown in Figure S5 in Supplement Material. The primary differences between the datasets arise from the selection of magnitude and distance ranges. In this study, the lower magnitude cutoff is set at M_{JMA} = 2.5, while BEA20 and DAW16 include only events with $M_{JMA} > 3$ and $M_{JMA} > 4$, respectively. In this flatfile we include all epicentral distances up to 1200 km, whereas DAW16 and BEA20 use epicentral distances ranging from 1 km and 0.6 km, respectively, up to 2000 km. Ad-

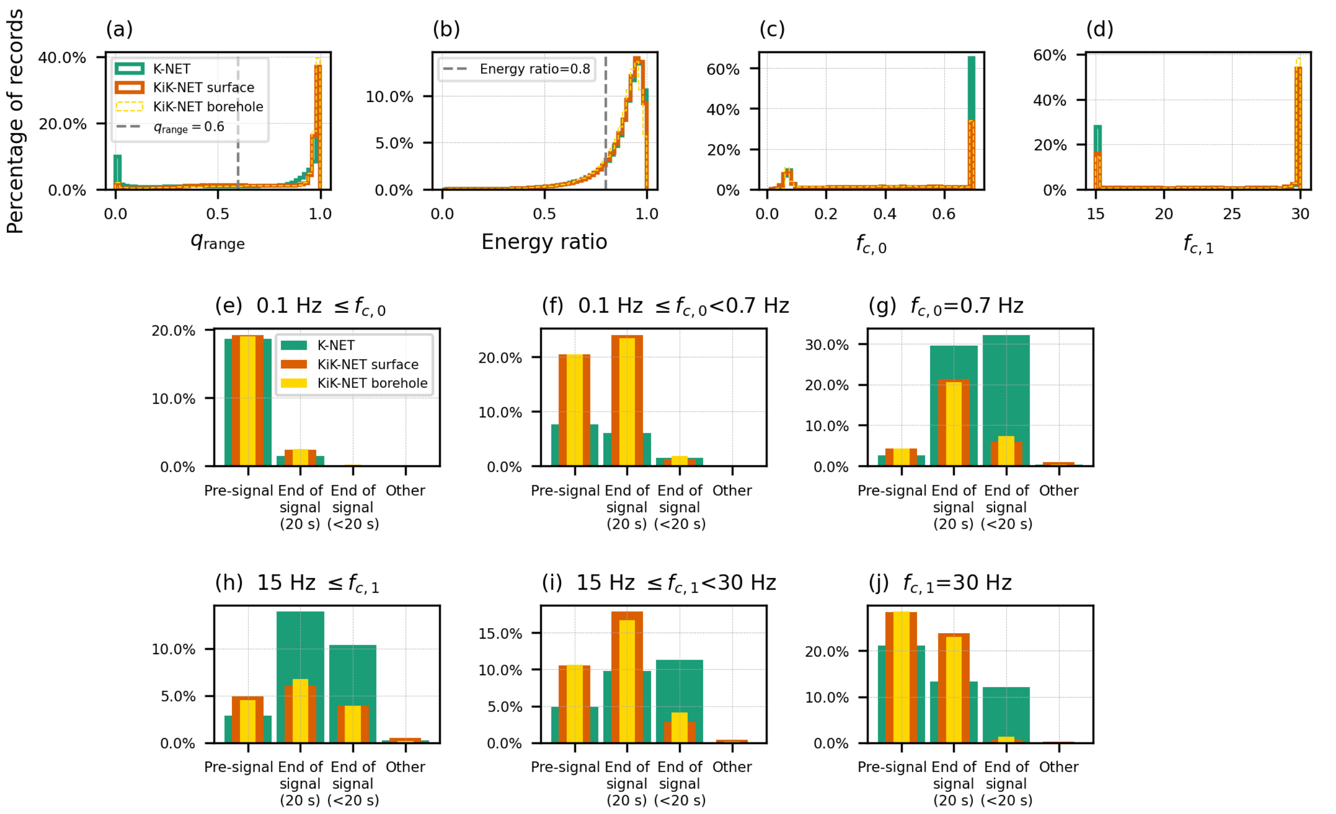

Figure 8 Distribution of key processing parameters for K-NET (green), KiK-net surface (orange), and KiK-net borehole (yellow). Top: (a) Usable frequency range ratio q_{range} (Equation 5). (b) Energy ratio of the main signal (P- and S-wave window) compared to the entire waveform. (c) Lower cutoff frequency for filtering. (d) Higher cutoff frequency for filtering. Middle and bottom: Distributions of the type of noise window used to select different ranges of the lower (middle) and higher (bottom) cutoff frequency ranges. The noise windows are separated into pre-event noise (prior to the P-arrival), and end of signal noise, either as exactly the last 20 seconds of the waveform, within the last 20 second of the waveform but with length shorter than 20 s, or sometime between the end of the signal and end of the waveform (other).

ditionally, a clear difference in the distribution of low PGA values between K-NET and KiK-net is observed, likely due to the different recording threshold levels for KiK-net and K-NET (Aoi et al., 2011).

An important processing difference between datasets is the method used for selecting cutoff frequencies in filtering. In NGA-West 2, cutoff frequencies were selected manually by visually checking the Fourier amplitude spectra and integrated displacement traces (Ancheta et al., 2014). DAW16, which was an early example of automatic processing, applied a set of predefined high-pass corner frequencies (0.07 Hz, 0.09 Hz, 0.14 Hz, 0.17 Hz, 0.22 Hz, 0.35 Hz, 0.46 Hz and 0.70 Hz) and used an iterative process to determine the most suitable frequency for each ground motion (Dawood et al., 2016).

Similar to this flatfile, BEA20 used the smoothed FAS of the noise and signal to select the corner frequencies. However, BEA20 selected the lower cutoff frequency as the point where the smoothed FAS of the entire waveform intersects the linearly scaled (by a factor of three) smoothed FAS of the noise, corresponding to SNR $=2$. For $f_{c,1}$, the noise was linearly scaled by a factor of two, corresponding to SNR $=1$ (Bahrampouri et al., 2020). Figure 9 shows the low (a) and high (b) corner frequencies for KiK-net from this study, NGA-West 2, DAW16, and BEA20. For the DAW16 dataset, the records were only high-pass filtered, so only the low corner frequencies for DAW16 are shown in figure 9. For both $f_{c,0}$ and $f_{c,1}$, the NGA-West2 values are in the lower range compared to the other datasets, as also reported by Dawood et al. (2016). However, since there are only four KiK-net events in the NGA-West 2 flatfile (corresponding to around 400 records), these values may not be representative of the entire KiK-net. For $f_{c,0}$, this study and BEA20 show an overall similar distribution, but with the BEA20 $f_{c,0}$ centered at slightly lower values, while our $f_{c,0}$ have more values set to maximum (0.7 Hz). This is likely because BEA20 consider SNR $=2$ and a slightly different method to select $f_{c,0}$. For $f_{c,1}$, the differences are larger, with $f_{c,1}$ ranging from 15 to 30 Hz for this flatfile, and from 25 to 100 Hz for BEA20. This is mainly due to the different maximum values used: here we set $f_{c,1,\text{max}} = 30\,\text{Hz}$, while for BEA20 $f_{c,1,\text{max}} = f_{\text{Nyquist}}\,\text{Hz}$. These differences are likely primarily due to the challenges associated with the limited availability of suitable noise windows for KiK-net and K-NET sites. However, this comparison also highlights the challenge of selecting cutoff frequencies for filtering in automatic processing of ground motions.

Within the engineering-relevant magnitude and distance ranges ($M_{\text{JMA}} \geq 5$ and $R_{\text{hypo}} < 100$), we compare PGA, PGV, and PGD for records common to this study

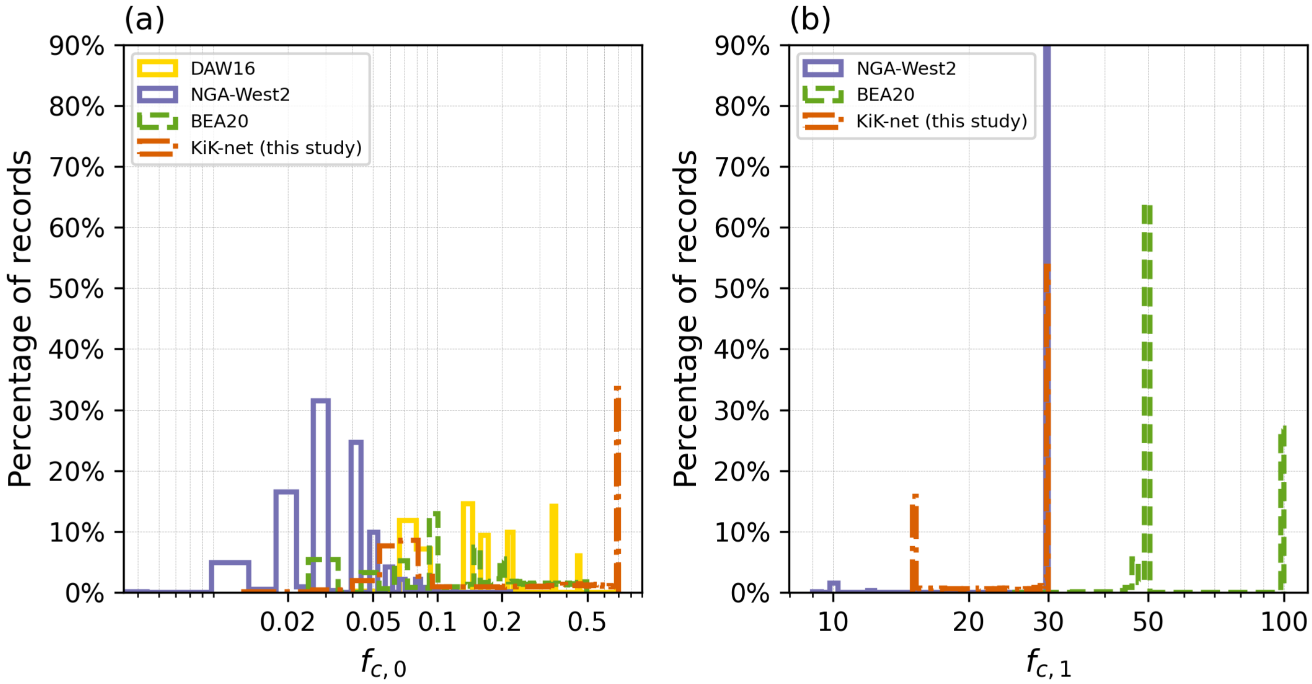

Figure 9 (a) Lower cutoff frequency for KiK-net (this study), NGA-West 2, DAW16, and BEA20, and (b) higher cutoff frequency for KiK-net (this study), NGA-West 2, and BEA20. Note that DAW16 only apply a high-pass filter and therefore do not include a higher corner frequency. This figure shows that the selection of cutoff frequencies for filtering is highly dependent on the processing and choice of the dataset designer.

(both KiK-net surface and K-NET) and NGA-West 2, and for records common to this study and DAW16 (Fig. 10). Note that because the BEA20 flatfile does not include PGV and PGD values, it is not included in this comparison. PGA has an overall good correlation between the different studies, with high Spearman's correlation coefficients ($\rho > 0.99$). For PGV, although the Spearman's correlation coefficients are still high ($\rho > 0.96$), there are also clear differences, particularly for low values and records with large absolute difference in selected $f_{c,0}$. This is also the case for PGD, which have the lowest correlation ($\rho = 0.622$) between KiK-net (surface) from NGA-West 2 and from DAW16. These differences emphasize the impact of processing procedures and cutoff frequency selection on PGV and, in particular, PGD values (e.g., Paolucci et al., 2011). While PGA is relatively insensitive to processing differences, PGV and PGD show greater variability depending on the filtering and baseline correction methodologies applied.

6 Summary and conclusions

The comprehensive and openly available strong-motion networks K-NET and KiK-net from NIED provide an invaluable resource for strong-motion seismology and an excellent opportunity to compile the available ground-motion data into a harmonized flatfile. Due to the large amount of data, the ground-motion records can realistically only be processed automatically. In this study we developed an automatic processing scheme for K-NET and KiK-net ground-motion records. The steps include matching the events with the JMA and F-net earthquake catalogs, P- and S-wave detection, SNR calculation, usable frequency range ratio estimation, filtering, baseline correction, and calculation of common intensity measures.

However, the large dataset also presents challenges that cannot be handled manually. In particular, the automatic selection of the usable frequency band and the cutoff frequencies for filtering is a key challenge that requires a sufficiently long noise window uncontaminated by earthquake signals. A suitable noise window is needed to accurately estimate the signal-to-noise ratio of the earthquake recordings. However, finding a sufficiently long and uncontaminated noise window is challenging due to the trigger-based recording method of Kyoshin networks. To address this, we propose a new method using the energy ratio of the noise window relative to the entire record to evaluate the quality of the available noise window. Finally, to avoid usable frequency bands that are too short, the cutoff frequencies are limited to a range between 0.01 and 0.7 Hz for the lower cutoff and between 15 and 30 Hz for the higher cutoff.

In a highly seismically active country like Japan, another challenge is that some of the records may contain multiple event arrivals that are difficult to detect automatically. Since the P- and S-wave windows of an earthquake are expected to contain most of the signal energy, we use the energy ratio between the P- and S-wave window and the entire waveform to assess whether additional event arrivals may be present outside the main window.

To evaluate the effect of the processing steps, a sys-

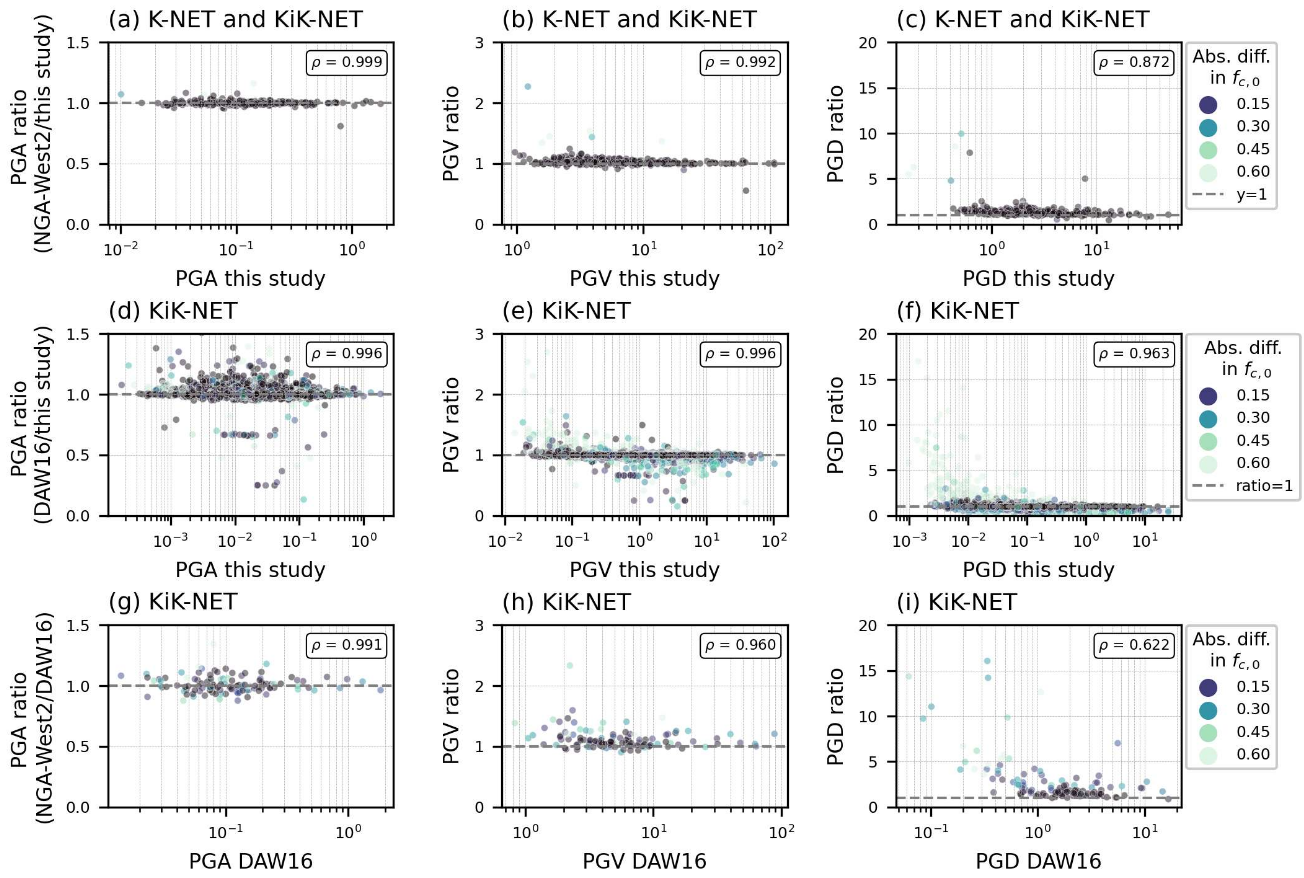

Figure 10 Comparison of peak ground-motion parameters across datasets. The left, middle, and right panels show peak ground acceleration (PGA), peak ground velocity (PGV), and peak ground displacement (PGD), respectively. The top row compares values from this study (KiK-net and K-NET) with NGA-West 2 (302 common values), the middle row compares KiK-net from this study with DAW16 (4,654 common values), and the bottom row compares DAW16 with KiK-net from NGA-West 2 (120 common values). Only strong motions in the engineering magnitude and distance range of interest ($M_{\mathrm{JMA}} \geq 5$ and $R_{\mathrm{hypo}} < 100$) are shown. Values are colored by the absolute difference in selected $f_{c,0}$ between datasets. Spearman's correlation coefficients (ρ) are provided in each figure.

tematic comparison with previously published flatfiles is performed. While the correlation coefficients are high ($\rho > 0.6$) for both the calculated rupture-based distances (R_{JB} and R_{rup}) and the peak ground amplitudes (PGA, PGV, and PGD), there are also clear differences. The differences are mainly due to the different calculation methods and, for PGV and PGD in particular, due to the difference in selected cut-off frequencies.

In total 950,796 ground motions from 18,738 earthquakes and 1,749 stations were processed, with 556,831 (59%) of these considered as high quality by having a usable noise window (noise energy ratio < 0.01), acceptable frequency range ratio ($q_{\mathrm{range}} > 0.6$), and low chance of multiple event contamination (signal energy ratio > 0.8). A ground motion flatfile is made available (see *Data and code availability* section) which includes metadata describing event and record properties, along with quality measures such as the usable frequency range ratio, SNR, and energy ratios. Additionally, the flatfile provides the intensity measures (peak amplitudes PGA, PGV, PGD), energy and duration measures (Arias Intensity, CAV, significant duration), RotD50 pseudo-spectral amplitudes, and Fourier amplitudes for all components. This flatfile is particularly suited for earthquake engineering research and systematic analyses of ground-motion properties, as it integrates recently recorded ground-motions from both KiK-net and K-NET, which have not previously been provided together in a flatfile of such large scale.

Acknowledgements

The authors would like to sincerely thank the National Research Institute for Earth Science and Disaster Prevention (NIED), Japan, for providing the data used in this article. The availability of these data is an irreplaceable and essential driving force for research in the engineering seismology community and for training of future researchers on a global scale. We also thank Marco Pilz and Dino Bindi, for helpful discussions and suggestions. We are grateful to the editor Dr. Pablo Heresi, an anonymous reviewer and Dr. Debi Kilb for helpful comments and suggestions. This research is funded by the European Commission, ITN-Marie Sklodowska-Curie New Challenges for Urban Engineering Seismology URBASIS-EU project, under Grant

Agreement 813137.

Data and code availability

The flatfile is published through the GFZ Data Services (Loviknes et al., 2026, http://doi.org/10.5880/GFZ.LKUT.2026.003). A description of the flatfile is provided in Supplementary Material 2. Future updated versions of the flatfile will also be made available at this Data Service.

The acceleration waveforms were downloaded from the Kyoshin-webpage (https://www.kyoshin.bosai.go.jp/, last accessed 2026-01-09). The JMA unified hypocenter catalog was downloaded from https://www.data.jma.go.jp/svd/eqev/data/bulletin/index_e.html (last accessed 2025-12-10), and the F-net catalog was downloaded from https://www.fnet.bosai.go.jp/event/mcata/index.php (last accessed 2026-01-21). The travel-time table from JMA was retrieved from https://www.data.jma.go.jp/svd/eqev/data/bulletin/catalog/appendix/trtime/tjma2001.zip (last accessed 2021-08-25).

Competing interests

The authors have no competing interests.

References

Akazawa, T. A technique for automatic detection of onset time of P-and S-phases in strong motion records. *13th World Conference on Earthquake Engineering, Vancouver, Canada*, 2004. https://www.iitk.ac.in/nicee/wcee/article/13_786.pdf.

Akkar, S., Sandıkkaya, M. A., Şenyurt, M., Azari Sisi, A., Ay, B. Ö., Traversa, P., Douglas, J., Cotton, F., Luzi, L., Hernandez, B., et al. Reference database for seismic ground-motion in Europe (RESORCE). *Bulletin of earthquake engineering*, 12:311–339, 2014. doi: 10.1007/s10518-013-9506-8.

Allen, R. V. Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, 68(5):1521–1532, 10 1978. doi: 10.1785/BSSA0680051521.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L. NGA-West2 Database. *Earthquake Spectra*, 30(3):989–1005, aug 2014. doi: 10.1193/070913EQS197M.

Aoi, S., Kunugi, T., and Fujiwara, H. STRONG-MOTION SEISMOGRAPH NETWORK OPERATED BY NIED: K-NET AND KiK-net. *Journal of Japan Association for Earthquake Engineering*, 4(3):65–74, 2004. doi: 10.5610/jaee.4.3_65.

Aoi, S., Kunugi, T., Nakamura, H., and Fujiwara, H. Deployment of new strong motion seismographs of K-NET and KiK-net. *Geotechnical, Geological and Earthquake Engineering*, 14:167–186, 2011. doi: 10.1007/978-94-007-0152-6_12.

Arias, A. A measure of earthquake intensity. *Seismic design for nuclear plants*, pages 438–483, 1970. https://www.osti.gov/biblio/4167721.

Bahrampouri, M., Rodriguez-Marek, A., Shahi, S., and Dawood, H. An updated database for ground motion parameters for KiK-net records. *Earthquake Spectra*, page 875529302095244, 9 2020. doi: 10.1177/8755293020952447.

Baker, J., Bradley, B., and Stafford, P. *Seismic hazard and risk analysis*. Cambridge University Press, 2021. doi: 10.1017/9781108425056.

Bastías, N. and Montalva, G. A. Chile Strong Ground Motion Flatfile. *Earthquake Spectra*, 32(4):2549–2566, 2016. doi: 10.1193/102715EQS158DP.

Bastías, N., Montalva, G., Leyton, F., Heresi, P., and Domínguez, H. An Extended and Updated Strong Ground-Motion Database for the Chilean Subduction Earthquakes. *Earthquake Spectra*, 42 (1):e70032, 2026. doi: 10.1002/esp4.70032.

Blaser, L., Krüger, F., Ohrnberger, M., and Scherbaum, F. Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bulletin of the Seismological Society of America*, 100(6):2914–2926, 2010. doi: 10.1785/0120100111.

Bommer, J. J. and Martínez-Pereira, A. The effective duration of earthquake strong motion. *Journal of Earthquake Engineering*, 03(02):127–172, 1999. doi: 10.1142/S1363246999000077.

Bommer, J. J., Stafford, P. J., and Alarcón, J. E. Empirical Equations for the Prediction of the Significant, Bracketed, and Uniform Duration of Earthquake Ground Motion. *Bulletin of the Seismological Society of America*, 99(6):3217–3233, 12 2009. doi: 10.1785/0120080298.

Bonilla, L. F. Y., Tsuda, K., Pulido, N., Regnier, J., and Laurendeau, A. Nonlinear site response evidence of K-NET and KiK-net records from the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, 63:785–789, 2011. doi: 10.5047/eps.2011.06.012.

Boore, D. M. On Pads and Filters: Processing Strong-Motion Data. *Bulletin of the Seismological Society of America*, 95(2):745–750, 04 2005. doi: 10.1785/0120040160.

Boore, D. M. Orientation-Independent, Nongeometric-Mean Measures of Seismic Intensity from Two Horizontal Components of Motion. *Bulletin of the Seismological Society of America*, 100(4):1830–1835, 08 2010. doi: 10.1785/0120090400.

Boore, D. M. and Atkinson, G. M. Boore-Atkinson NGA Ground Motion Relations for the Geometric Mean Horizontal Component of Peak and Spectral Ground Motion Parameters. *PEER Report 2007-01. Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA*, May 2007. https://peer.berkeley.edu/publications/2007-01.

Boore, D. M. and Bommer, J. J. Processing of strong-motion accelerograms: needs, options and consequences. *Soil Dynamics and Earthquake Engineering*, 25(2):93–115, 2005. doi: 10.1016/j.soildyn.2004.10.007.

Boore, D. M. and Goulet, C. A. The effect of sampling rate and anti-aliasing filters on high-frequency response spectra. *Bulletin of earthquake engineering*, 12:203–216, 2014. doi: 10.1007/s10518-013-9574-9.

Boore, D. M., Watson-Lamprey, J., and Abrahamson, N. A. Orientation-Independent Measures of Ground Motion. *Bulletin of the Seismological Society of America*, 96(4A):1502–1511, 08 2006. doi: 10.1785/0120050209.

Boore, D. M., Azari Sisi, A., and Akkar, S. Using Pad-Stripped Acausally Filtered Strong-Motion Data. *Bulletin of the Seismological Society of America*, 102(2):751–760, 04 2012. doi: 10.1785/0120110222.

Brune, J. N. Tectonic stress and the spectra of seismic shear waves from earthquakes. *Journal of geophysical research*, 75 (26):4997–5009, 1970. doi: 10.1029/JB075i026p04997.

Buscetti, M., Traversa, P., Perron, V., Rischette, P., Hollender, F., and Shible, H. Epos-France Database of Earthquake Ground Motion for Mainland France: 1996–2021 Updated Version. *Seismologi-*

cal Research Letters, 09 2024. doi: 10.1785/0220240053.

Castro, S., Benavente, R., Crempien, J. G. F., Candia, G., and de la Llera, J. C. A Consistently Processed Strong-Motion Database for Chilean Earthquakes. *Seismological Research Letters*, 93(5): 2700–2718, 06 2022. doi: 10.1785/0220200336.

Chiou, B. S. and Youngs, R. R. NGA Model for Average Horizontal Component of Peak Ground Motion and Response Spectra. *PEER Report 2008-09. Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA*, November 2008.

Dawood, H. M. and Rodriguez-Marek, A. A Method for Including Path Effects in Ground-Motion Prediction Equations: An Example Using the Mw9.0 Tohoku Earthquake Aftershocks. *Bulletin of the Seismological Society of America*, 103(2B):1360–1372, 05 2013. doi: 10.1785/0120120125.

Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E. A flatfile for the KiK-net database processed using an automated protocol. *Earthquake Spectra*, 32(2):1281–1302, may 2016. doi: 10.1193/071214EQS106.

Derras, B., Bard, P. Y., Régnier, J., and Cadet, H. Non-linear modulation of site response: Sensitivity to various surface ground-motion intensity measures and site-condition proxies using a neural network approach. *Engineering Geology*, 269:105500, 5 2020. doi: 10.1016/J.ENGGEO.2020.105500.

Esfahani, R. D. D., Cotton, F., Ohrnberger, M., and Scherbaum, F. TFCGAN: Nonstationary Ground-Motion Simulation in the Time–Frequency Domain Using Conditional Generative Adversarial Network (CGAN) and Phase Retrieval Methods. *Bulletin of the Seismological Society of America*, 113(1):453–467, 12 2022. doi: 10.1785/0120220068.

Fujii, Y. and Satake, K. Slip distribution of the 2024 Noto Peninsula earthquake (MJMA 7.6) estimated from tsunami waveforms and GNSS data. *Earth, Planets and Space*, 76:44, 2024. doi: 10.1186/s40623-024-01991-z.

Fukuyama, E. Automated seismic moment tensor determination by using on-line broadband seismic waveforms [in Japanese with English abstract]. *J. Seismol. Soc. Jpn.*, 51:149, 1998. doi: 10.4294/zisin1948.51.1_149.

Hastie, T., Tibshirani, R., and Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009. doi: 10.1007/978-0-387-84858-7.

Havskov, J. and Ottemöller, L. *Signal Processing*. Springer Netherlands, Dordrecht, 2010. doi: 10.1007/978-90-481-8697-6_4.

Ito, Y., Yamazaki, S., and Kurahashi, T. Geological features of landslides caused by the 2018 Hokkaido Eastern Iburi Earthquake in Japan. *Geological Society, London, Special Publications*, 501(1): 171–183, 2021. doi: 10.1144/SP501-2019-122.

Itoh, Y. Coseismic Slip and Early Afterslip of the 2024 Hyuganada Earthquake Modulated by a Subducted Seamount. *Geophysical Research Letters*, 52(2):e2024GL112826, 2025. doi: 10.1029/2024GL112826.

Kaklamanos, J., Bradley, B. A., Thompson, E. M., and Baise, L. G. Critical parameters affecting bias and variability in site-response analyses using KiK-net downhole array data. *Bulletin of the Seismological Society of America*, 103(3):1733–1749, 2013. doi: 10.1785/0120120166.

Kanamori, H. The energy release in great earthquakes. *Journal of Geophysical Research (1896-1977)*, 82(20):2981–2987, 1977. doi: 10.1029/JB082i020p02981.

Kobayashi, H., Watanabe, T., and Koketsu, K. Rupture processes of the 2021 and 2022 Fukushima-oki earthquakes: adjacent events on the complex fault system in the subducting slab. *Earth, Planets and Space*, 75:81, 2023. doi: 10.1186/s40623-023-01838-z.

Kokusho, T. and Sato, K. Surface-to-base amplification evaluated from KiK-net vertical array strong motion records. *Soil Dynamics and Earthquake Engineering*, 28(9):707–716, 2008. doi: 10.1785/BSSA0860061733.

Lanzano, G., Sgobba, S., Luzi, L., Puglia, R., Pacor, F., Felicetta, C., D'Amico, M., Cotton, F., and Bindi, D. The pan-European Engineering Strong Motion (ESM) flatfile: compilation criteria and data statistics. *Bulletin of Earthquake Engineering*, 17:561–582, 2019. doi: 10.1007/s10518-018-0480-z.

Lilienkamp, H., von Specht, S., Weatherill, G., Caire, G., and Cotton, F. Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach. *Bulletin of the Seismological Society of America*, 112(3):1565–1582, 03 2022. doi: 10.1785/0120220008.

Loviknes, K., von Specht, S., Lilienkamp, H., Händel, A., Zhu, C., and Cotton, F. KiK-net and K-NET flatfile with automatically processed ground motions and associated metadata. V. 0002 [dataset], 2026. doi: 10.5880/GFZ.LKUT.2026.003.

Luzi, L., Lanzano, G., Felicetta, C., D'Amico, M., Russo, E., Sgobba, S., and Pacor, F. ORFEUS Working Group 5: Engineering Strong Motion Database (ESM)(Version 2.0), Istituto Nazionale di Geofisica e Vulcanologia (INGV), 2020. doi: 10.13127/ESM.2.

Manea, E. F., Bora, S. S., Hutchinson, J. A., and Kaiser, A. E. Uniformly Processed Fourier Spectra Amplitude Database for Recently Compiled New Zealand Strong Ground Motions. *Seismological Research Letters*, 95(1):239–252, 12 2023. doi: 10.1785/0220230228.

Morikawa, N., Iwaki, A., Fujiwara, H., Akiyama, S., Maeda, T., Kubo, H., Aoi, S., Hayakawa, T., Takahashi, M., Kato, K., Tomozawa, Y., Suzuki, F., Motoki, K., Iida, T., Sato, T., Ishii, T., Miyakoshi, J., Shimazu, N., Oana, A., Hayashi, T., Kishida, N., T., O., Si, H., Matsuyama, H., and Midorikawa, S. Flat file of K-NET and KiK-net strong-motion records. *Proceedings in the 18th World Conference on Earthquake Engineering*, 2024.

National Research Institute for Earth Science and Disaster Resilience. Digital data of source process for large earthquakes based on strong motion records. *National Research Institute for Earth Science and Disaster Resilience.*, 2025. doi: 10.17598/NIED.0031.

National Research Institute for Earth Science and Disaster Resilience (NIED). NIED K-NET, KiK-net. *National Research Institute for Earth Science and Disaster Resilience*, 2019. doi: 10.17598/NIED.0004.

Nigam, N. C. and Jennings, P. C. *Digital calculation of response spectra from strong-motion earthquake records*. Jan 1968.

Nigam, N. C. and Jennings, P. C. Calculation of response spectra from strong-motion earthquake records. *Bulletin of the Seismological Society of America*, 59(2):909–922, 04 1969. doi: 10.1785/BSSA0590020909.

Okada, Y., Kasahara, K., Hori, S., Obara, K., Sekiguchi, S., Fujiwara, H., and Yamamoto, A. Recent progress of seismic observation networks in Japan - Hi-net, F-net, K-net and KiK-net -. *Earth, Planets and Space*, 56:xv–xxviii, 2004. doi: 10.1186/BF03353076.

Paolucci, R., Pacor, F., Puglia, R., Ameri, G., Cauzzi, C., and Massa, M. Record processing in ITACA, the new Italian strong-motion database. *Earthquake Data in Engineering Seismology: Predictive Models, Data Management and Networks*, pages 99–113, 2011. doi: 10.1007/978-94-007-0152-6_8.

Perron, V., Laurendeau, A., Hollender, F., Bard, P. Y., Gélis, C., Traversa, P., and Drouet, S. Selecting time windows of seismic phases and noise for engineering seismology applications: a versatile methodology and algorithm. *Bulletin of Earthquake Engineering*, 16:2211–2225, 6 2018. doi: 10.1007/s10518-017-0131-9.

Roten, D. and Olsen, K. B. Estimation of Site Amplification from

Geotechnical Array Data Using Neural Networks. *Bulletin of the Seismological Society of America*, 111(4):1784–1794, 03 2021. doi: 10.1785/0120200346.

Régnier, J., Cadet, H., Bonilla, L. F., Bertrand, E., and Semblat, J. F. Assessing nonlinear behavior of soils in seismic site response: Statistical analysis on KiK-net strong-motion data. *Bulletin of the Seismological Society of America*, 103:1750–1770, 6 2013. doi: 10.1785/0120120240.

Sandıkkaya, M. A., Güryuva, B., Özkan Kale, Okçu, O., İçen, A., Yenier, E., and Akkar, S. An updated strong-motion database of Türkiye (SMD-TR). *Earthquake Spectra*, 40(1):847–870, 2024. doi: 10.1177/87552930231208158.

Scherbaum, F., Schmedes, J., and Cotton, F. On the Conversion of Source-to-Site Distance Measures for Extended Earthquake Source Models. *Bulletin of the Seismological Society of America*, 94(3):1053–1069, 06 2004. doi: 10.1785/0120030055.

Shannon, C. E. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, 1949. doi: 10.1109/JRPROC.1949.232969.

Thingbaijam, K. K. S., Martin Mai, P., and Goda, K. New Empirical Earthquake Source-Scaling Laws. *Bulletin of the Seismological Society of America*, 107(5):2225–2246, 09 2017. doi: 10.1785/0120170017.

Traversa, P., Maufroy, E., Hollender, F., Perron, V., Bremaud, V., Shible, H., Drouet, S., Guéguen, P., Langlais, M., Wolyniec, D., Péquegnat, C., and Douste-Bacque, I. RESIF RAP and RLBP Dataset of Earthquake Ground Motion in Mainland France. *Seismological Research Letters*, 91(4):2409–2424, 05 2020. doi: 10.1785/0220190367.

Wang, Z., Sun, J., Ito, E., Kawase, H., and Matsushima, S. Empirical method for deriving horizontal site amplification factors considering nonlinear soil behaviors based on K-NET and KiK-net records throughout Japan. *Bulletin of Earthquake Engineering*, pages 1–26, 2023. doi: 10.1007/s10518-023-01712-z.

Wells, D. L. and Coppersmith, K. J. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, 84(4):974–1002, 08 1994. doi: 10.1785/BSSA0840040974.

Yoshida, K., Uno, M., and Bassett, D. The 2025 Mw7.6 Aomori-Oki Megathrust Sequence and a Slip-Parallel Seismic Belt to the Trench. *ESS Open Archive*, 2026(0222), 2026. doi: 10.22541/essoar.177177384.41366223/v1.

Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., Fukushima, Y., and Fukushima, Y. Attenuation Relations of Strong Ground Motion in Japan Using Site Classification Based on Predominant Period. *Bulletin of the Seismological Society of America*, 96(3):898–913, 06 2006. doi: 10.1785/0120050122.

Zhu, C., Weatherill, G., Cotton, F., Pilz, M., Kwak, D. Y., and Kawase, H. An open-source site database of strong-motion stations in Japan: K-NET and KiK-net (v1. 0.0). *Earthquake Spectra*, 37(3):2126–2149, 2021. doi: 10.1177/8755293020988028.