Letter to Editor

Dear Editor Pablo Heresi, thank your for comments and for your patience. Following the comments of the reviewers, and due to new data and some errors in the processing algorithm, we have now updated the flatfile and revised the manuscript. In addition, we have added a new author: Chuanbin Zhu.

The main changes to the flatfile and manuscript are:

- New ground motion data; up until the end of 2025
- Change in distance computation; correction to R_{hypo} calculation and R_{JB} and R_{rrup} are now computed using NIED rupture models when available.
- Change of title: “Harmonization of KiK-net and K-NET Ground-Motion Parameters and Metadata: A Flatfile of 950,000 Three-Component Records Spanning 29 Years”
- We add the data description file as Supplementary Material

A more detailed list of changes to the processing and the flatfile is listed at the end of this document. The more detailed response to the reviewers are given below.

We are at the moment still waiting for the new DOI, the updated flatfile is therefore for the moment available at: https://nextcloud.gfz.de/s/9Tfy8m2B5zjwPW3.

Reviewer A

Line #20-21. I recommend including the DOI link to the repository in the abstract to facilitate quick and easy access for readers and potential users of flatfile.

- Thank you for the suggestion, we have added this.

Line #31. In the context of the article’s main scope - developed and published a flatfile summarizing waveform information and their intensities – reference could also be made to Bastias and Montalva (2016), who pursued a similar scope.

- Thank you for the suggestion, we have added this reference in the introduction.

Line #47-49. This sentence requires clarification. Is this study an extension of DAW16 or BEA20, or does it provide new features, intensities, or a new/different processing scheme? If the study can be considered as only an extension of prior flatfile this should be explicitly stated in the abstract.

- Thank you for pointing out this lack of clarity, our flatfile is independent of the DAW16 and BEA20, and differs in both processing steps and included variables, We have rephrased in order to make this clearer. In addition to the new records, the main differences between our flatfile and DAW16 and BEA20 are:

- This flatfile includes K-Net (both DAW16 and BEA20 only includes KiK-Net)
- Different methods for selection cutoff frequencies
- Different methods for evaluation the quality of the records; e.g. multiple arrival detection. In this study we use the energy ratio criteria.
- This flatfile includes additional intensity measures, including Arias Intensity, cumulative absolute velocity, rotD50, rotD00, and rotD100 PGA, PGV, PGD and response spectra

Line #70-71. What is the impact of missing Vs30 available on the data set? How many measured Vs30 are available (both total number and percentage of the total).

- All site parameters shown in the manuscript are retrieved from the site database by Zhu et al 2021. In the original flatfile, none of these values were included. In the new updated flatfile we include the VS values for surface and borehole, VS10, VS30, depth to 800 m/s and 1500 m/s and resonance frequency f0 from Zhu et al. (2021). Zhu et al derives the VS values from 1D velocity profiles published by NIED, and while the KiK-net boreholes reach at least 100 m, the velocity profiles for K-NET (where there are no borehole stations) only reach down to 20 meters. For this reason most K-NET sites does not have valid VS30 values.

Line #71-72. A brief description of the Zhu et al. (2021) should be included. Site database is a crucial part of a ground motion flatfile.

- We have added a paragraph describing the the Zhu et al (2021) dataset.

Line #73-80. This paragraph is duplicated. I recommend reformulated, rephrasing and/or consolidating it.

- Thank you for pointing this out, we have removed the duplicated paragraph.

Figure 1. Add x- and y-labels to panels (c) and (d)

- Thank you for pointing this out, we have added the labels

Line 91-92. In the merging process discusses magnitude and time windows, but what about distance windows? For example, Weatherill et al. (2016) used a time-window screening of 60 seconds and a distance-window of 110 km to classify two events as duplicated. Could a distance windows (hypocentral or epicentral) also be applied?

- Because the magnitude range considered is large and some events have long rupture lengths, using epicentral or hypocentral distance as criteria would not be appropriate. Instead we are using the estimated surface rupture length as a distance criteria; events are matched when their epicenters are within 3 times the surface rupture length. We have rewritten the paragraph to make this clearer.

Line 92-93. How does the unreliability of depth information affect the dataset overall? How many records are assigned a standard Depth? Is a significant part of the dataset?

- Out of the 20,742 events recorded by KiK-net and Knet, 320 events have depth=0. The depth uncertainty therefore does not have a major effect on the dataset. This information has been added to the manuscript.

Line 99-100. Why were unmatched records excluded? If these are real earthquake recordings, extra valuable information could be lost.

- Matching with JMA or Fnet is important because the original metadata origin times are only accurate down to the minute. When determining whether the record includes the P- or S-wave arrivals, the accurate origin time is needed. The events without this information (so without a JMA or FNET match) are therefore excluded. The paragraph have been rephrased for more clarity on this. Also, as shown in Figure S1, about 90% of the events (18,782 out of 20,742) recorded by KiK-net and Knet are matched with either JMA and Fnet. The majority of the unmatched events are either low magnitude or only recorded at far distances. Our goal is not to reproduce the NIED database but extract from the database the records which can be documented with high quality metadata.

Line 100-101. Are magnitude scales available for all events, or in some cases? Is there a priority list of magnitudes (e.g., preferring moment magnitude)? In case of preferred magnitude is not available, is a homologated equation to convert to the priority magnitude used to estimate it?

- The JMA magnitude particular for Japan is available for all events (column name: *Mag._Meta*), while the JMA and Fnet catalog includes additional magnitudes. A priority list of magnitudes is user-dependent and therefore not included in this manuscript.

Figure 2. add x and y-label to panels (b) through (e).

- Thank you for pointing this out, we have added the labels

Line 114. Why were only events in F-net catalog used? The paper should further justify this preference. For example, was it due to missing moment tensor solutions in other catalogs? Why not use simulated rupture planes based on hypocentral data (e.g., Mai et al., 2005)?

- The F-net catalog is the most complete event catalog for Japan and is designed for Japan, we therefore mainly use the F-net catalog in this dataset. In the new updated flatfile, for the larger events and when available, we also use rupture planes from source inversions published by NIED are to compute the RJB and rupture distances.

Line 117-118. This methodology section is unclear and should be rephrased. Is the centroid obtained from F-net catalog or another source? While F-net provides moment tensors, does it also provide centroid locations? Both parameters are related but it is not clear if both information is provided with the same reference.

- Both the centroid and the moment tensors used to estimate the rupture planes are from the Fnet catalog. We have rephrased the sentence to make this clearer.

Figure 3. I expected that in the marginal plots (or histograms) the y-label to be "counts" or "frequency" but is necessary complete to a clear understanding of the figure. Please complete this for clarity.

- Thank you for pointing this out, we have added the labels

Line 154-155. How is it determined that a record starts after the S-wave arrival? A brief explanation would be helpful

- The P- and S-wave arrivals are determined using a travel-time table from JMA, the paragraph is rephrased to make this clearer.

Line 157-158. If travel-time tables depend on focal depth and epicentral distance, are we referring to hipocentral distance? Please clarify the distinction and provide the expression used.

- The JMA travel-time table provides the P- and S-wave travel times as a function of focal depths (range: 0 - 700 km) and epicentral distance (range: 0 - 2000 km), bilinear interpolation is then used to obtain any value within the table's range. Using hypocentral distance and depth would therefore mean to count depth twice. We have rewritten the sentence to make this clearer.

Line 158-162. This paragraph is confusing and difficult to follow. I recommend reformulating it for clarity.

- Thank you for pointing this out, we have rewritten the paragraph.

Line 167-168. Add a reference(s) to support this condition

- We have added the Perron et al., 2018 article as reference.

Figure 4. What is the meaning of the colors in the flowchart? Also, the "output" box appears to be disconnected from the flow.

- Thank you for pointing out that this description is missing. We have added a description of the signification of the colors.

Line 179. Why Ds_min is 5 seconds and Ds_max is 120 seconds. Please justify these thresholds.

- DSmin is set to 5s to ensure suffient spectral resolution at low frequencies particularly for waveforms recorded at stations close to the hypocenter (low TS-TP). Dsmax is set to 120s as this is the minimum length of the raw waveforms from the Kyoshin networks (see Figure 7 and Okada et al., 2004). We have added a justification for these values.

Line 181-183. The two consecutive sentences are very similar. Consider deleting one or rewriting both for conciseness.

- Thank you for pointing this out, we have reformulated the sentences.

Figure 5. Is the blue energy trace also behind the orange line? If so, add an explanatory note to the figure caption.

- Yes the blue energy trace is behind the orange, this has been clarified in the caption of the figure.

Line 189-190. Does the P- and S-wave arrival, based on STA/LTA, really work in all the cases and traces analyzed? In practice, automatic detection of P-wave arrivals is often successful, but identifying S-wave arrivals is more challenging, particularly in noisy records. How was this handled in your study?

- Yes, identifying the S-wave is more challenging, we therefore only detect the first arrival with this method. Meaning that as long as the P-wave is within the recorded waveform,we detect the P-wave arrival. If not, we detect the S-wave arrival. We have added a clarifying sentence.

Line 206-207. Please add the original reference for the Arias Intensity (Arias, 1970).

- Thank you for pointing out this missing reference, it has been added.

Line 225. Why was the threshold of SNR>3 chosen? Add a reference or a justification of this selection.

- We have added the Boore and Boomer, 2005, paper as reference

Line 231-232. Why is analysis done per component, but processing uses the most conservative component? This approach might discard valuable data from components with high SNR that are filtered out due to a low-SNR orthogonal component. Why not analyze and process each component independently?

- We use the same cutoff frequencies for all components because in earthquake engineering, where flatfiles are much used, it is common to use the most conservative estimate and combine the components for ground motion analysis (e.g., Boore, 2005; Ancheta et al, 2014) . We are also using the most conservative of the components due to the issue of finding a long and high quality noise info for the these frequency analysis.

Line 253-254. This justification could be improved. Why oversampling? Oversample the signal is generally not recommended under the Nyquist–Shannon theorem. What are the actual differences in computed intensities between raw and oversampled traces?

- The purpose of oversampling is to estimate the largest potential extrema within the dataset, this is also done by Nigam and Jennings, 1969 which is both the methods used here and the most common method for computing response spectra in earthquake engineering. Boore and Goulet, 2014, among others, discussed that linear interpolation of the time-series could lead to underestimation of the absolute peaks and introduce high frequency noise above the Nyquist frequency, and suggested the sinc interpolation as an alternative. The sinc interpolation avoids these issues by zero-padding the Fourier spectra beyond before transforming back to time domain. We have added a sentence to improve the justification.

Line 263-265. What is the step size of the provided increments? Please add this to the manuscript. Also, regarding the rotated-intensity analysis: are sensor orientations presented in the flatfile? Or are all stations oriented to real north?

- The increments are 1 degree each, this has been added to the manuscript. The sensor orientation for the KiK-net and Knet stations are related to much uncertainty and not considered in the flatfile, however, as also described by Boore (2010) the rotDxx measures are orientation independent.

Line 301-302. How is the signal end defined? Is it the end of the S-wave? Please clarify.

- The end of the main signal is determined using the Perron et al., (2018) methods based on the Brune's spectra, as described in the Processing steps section

Figure 8. Panels (c) and (d) show lower and upper cutoff frequencies near the minimum and maximum preselected values (Table 2), while in Figure 9 shows more uniform distributions across frequencies. What is the difference between both figures?

- In the original manuscript the two figures used Different y-axis (% in Fig.8 while log in Fig.9) which made the impression that the distributions were different. To avoid this, we have now change the y-axis of Fig.9 to match Fig.8.

Line 369-372. This sentence comparing flatfiles is unclear and not very informative. I recommend using a consistent distance metric across comparisons to improve clarity.

- Thank you for the suggestion, we have changed the sentence and the cutoff criteria to using epicenter distance to ensure consistency.

Figure 10. In panel (d), some records deviate notably from the 1.0 ratio line ("DAW16–this study"), and these seem to match outliers in PGV (panel e). Why are no outliers shown in panel (g) comparing NGA–DAW16 (similar to panel (a))? Do these outlier records share any characteristics or processing conditions?

- The outliers in figure 10d and e are related to records that are not included in the NGA-West 2 flatfile (which only has 120 common KiK-net records with Dawood) and the outliers are therefore not shown in figure10g. The large difference in PGA and PGV values between this study and Dawood shown as outliers in figure 10d and e are mainly related to difference in lower cutoff frequency (fc0). The scatterpoints in figure10 are colored by absolute difference in fc0, however the difference between low fc0 is not as evident from the plot. To further demonstrate that the outliers (here defined as PGA and PGV ratio < 0.8, amounting to 34 and 69 records) are cased by the difference in fc0, we show here the distribution of fc0 for the two datasets for the outlier records.

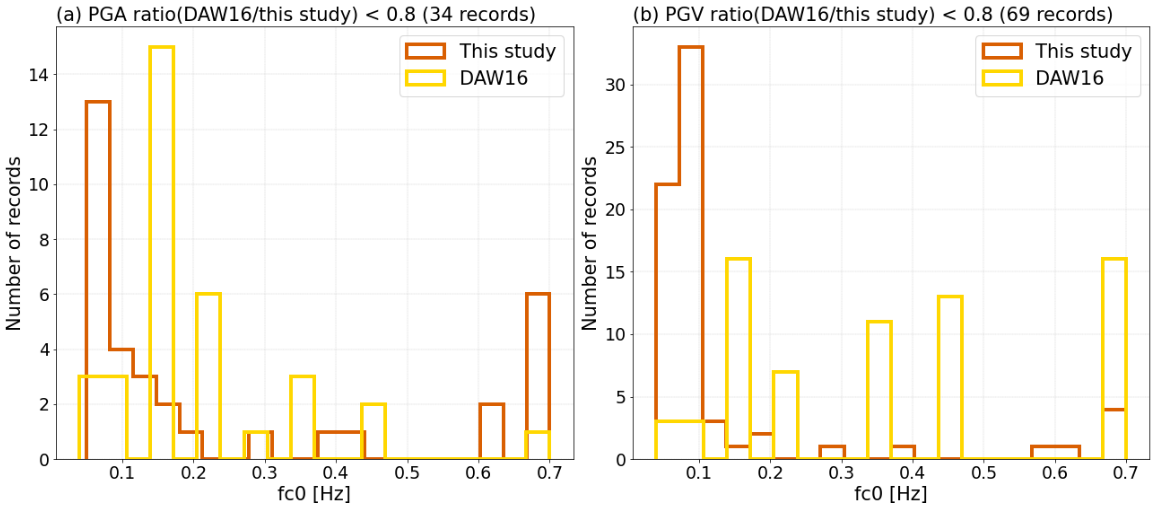

Figure 1: Difference in distribution of fc0 for (a) PGA ratio (DAW16/this study) < 0.8, and (b) PGV ratio <0.8.

Line 436. Why is only RotD50 computed? Why were the extreme values of the distribution (RotD0 or RotD100) excluded?

- Thank you for the suggestion, in the new version of the flatfile we are including the RotD0 and RotD100 as well.

References

Arias A. A measure of earthquake intensity, In Seismic Design for Nuclear Power Plants, Hansen RJ (ed.). MIT Press: Cambridge, MA, 1970; 438 – 483.

Bastías N, Montalva GA. Chile Strong Ground Motion Flatfile. Earthquake Spectra. 2016;32(4):2549-2566. doi:10.1193/102715EQS158DP

P. Martin Mai, P. Spudich, J. Boatwright; Hypocenter Locations in Finite-Source Rupture Models. *Bulletin of the Seismological Society of America* 2005;; 95 (3): 965–980. doi: https://doi.org/10.1785/0120040111

Weatherill GA, Pagani M and Garcia J (2016) Exploring earthquake databases for the creation of magnitude-homogeneous catalogues: tools for application on a regional and global scale. Geophysical Journal International 206(3): 1652–1676.

New references:

T. D. Ancheta, R. B. Darragh, J. P. Stewart, E. Seyhan, W. J. Silva, B. S.-J. Chiou, K. E. Wooddell, R. W. Graves, A. R. Kottke, D. M. Boore, T. Kishida, and J. L. Donahue. NGA-West2 Database. *Earthquake Spectra*, 30(3):989–1005, aug 2014. ISSN 8755-2930. doi: 10.1193/070913EQS197M.

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

D. M. Boore. 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. ISSN 0037-1106. doi: 10.1785/0120090400.

N. C. Nigam and P. C. Jennings. Calculation of response spectra from strong-motion earthquake records. *Bulletin of the Seismological Society of America*, 59(2):909–922, 04 1969. ISSN 0037-1106. doi: 10.1785/BSSA0590020909.

V. Perron, A. Laurendeau, F. Hollender, P. Y. Bard, C. Gélis, P. Traversa, and S. Drouet. 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. ISSN 15731456. doi: 10.1007/s10518-017-0131-9.

Y. Okada, K. Kasahara, S. Hori, K. Obara, S. Sekiguchi, H. Fujiwara, and A. Yamamoto. 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. ISSN 18805981. doi: 10.1186/BF03353076.

C. Zhu, G. Weatherill, F. Cotton, M. Pilz, D. Y. Kwak, and H. Kawase. 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.

Reviewer E

I was able to navigate to the data portal (see snapshot below), but I was not able to access the data. When I attempted to click on the links, I got the error “File not found” for both the dataset and the data description pdf file.

- We apologize for this error, however are not sure why it occurred as it seem to work now. For the new updated dataset we have added a link to our email for directly sharing of the data in case of technical issues like this.

Index of /download/10.5880.GFZ.LKUT.2025.001-CVaecnui

Icon	Name	Last modified	Size	Description
[PARENTDIR]	Parent Directory		-	
[DIR]	2025-001_Loviknes-et-al_data/	2025-02-26 14:24	-	
[ ]	2025-001_Loviknes-et-al_Data-Description.pdf	2025-02-26 08:51	243K	

The authors might consider these suggestions to strengthen the manuscript:

Consider revamping your title to make it clear that through this work you are providing an extensive ground motion database for future use by other researchers.

OLD: Harmonized KiK-net and K-NET flatfile for systematic analysis of earthquake ground motions

NEW: Compiled dataset of ~900,000 ground motions (~28 years; 3-component measurements) of KiK-net (surface stations) and K-NET (surface and borehole stations) and associated metadata

- Thank you for the suggestion, we have changed the title to "Harmonization of KiK-net and K-NET Ground-Motion Parameters and Metadata: A Flatfile of 950,000 Three-Component Records Spanning 29 Years"

L217-218: I was surprised by this decision: "If the observed P-wave arrival occurs before the record start, the signal is cut to begin 5 seconds before the S-wave arrival." Why aren't these data discarded? Consider adding a sentence about why this type of data might be useful to someone else down the line.

- For strong ground-motion analysis the S-wave window holds the most energy and is therefore the most interesting. Previous flatfiles (e.g. Bharampouri et al. 2020) have used only the S-wave window of the ground motions when deriving their intensity measures. While we in this dataset retain the whole waveform from the P-wave when possible we also consider the waveforms that only include S-wave as useful and include them. The information on whether the record starts at P- or S-wave arrival is included in the flatfile under the *WaveType* column.

In the supplement, consider adding a table listing the flatfile parameter values, a column with a description of what these parameters are, and an additional column listing how these data might be best quality controlled so potentially bad data can be identified.

- Thank you for the suggestion we now include the data description that is published along the dataset as a supplement to this manuscript.

Consider adding a sentence to the conclusion about how you plan to update these data – will this be done yearly? Where will these new data be located?

- We hope to provide an updated flatfile every 1-2 years published as updated flatfiles on the same data service center (GFZ). We have added a sentence about this in the Data and code availability section.

I list additional minor suggestions below, listed by line number.

I look forward to seeing this manuscript in print.

Dr. Debi Kilb (OK to release my name to the authors)

L131: Consider adding an additional sentence explaining why you prefer using the method of Blaser et al. (2010), over Wells and Coppersmith (1994).

- Thank you for the suggestion, we have added a sentence explaining this. Blaser et al. (2010) is based on a larger dataset than Wells and Coppersmith and also includes subduction events.

L144-147: I agree with the authors that the comparisons that they are making, and the overall consistency of their results with other studies (Spearman correlation coefficients $\rho>0.96$) supports that the results are as expected. They point out the mismatch in distances for some of the larger events, and give what, to my mind, is reasonable reasons for the discrepancies. No red flag about their method used here on my end.

- Thank you.

L152-161: Nice explanation of how you use the triggered data.

- Thank you.

L166-170: I'm very interested in this new energy ratio parameter, great idea to figure out how to derive it and to include it!

- Thank you.

L217-218: I was surprised by this decision: "If the observed P-wave arrival occurs before the record start, the signal is cut to begin 5 seconds before the S-wave arrival." Why aren't these data discarded? Consider adding a sentence about why this type of data might be useful to someone else down the line.

- See reply given above.

L231: Nice! “Each corner frequency is determined separately for the horizontal components, and the more conservative of the two is selected”.

- Thank you

L236: You might note here that all information is included in the flat file and suspect data removal is left to the database user, although you do provide suggested guidelines to identify bad data.

- Thank you for the suggestion, we have added a sentence stating this.

L253-254: I expect just a typo, but you should avoid a one-sentence paragraph. Combine this text with the next paragraph.

- Thank you for pointing this out, the sentence have been combined with the next paragraph.

L275: Consider adding what percent of these data were successfully processed – 91% I think, double check my math.

- Thank you for the suggestion, we have added this.

L275-284: I like that you took the time to explain why some of these data were rejected.

- Thank you

L299: Consider adding what percentage of the data this represents.

- Thank you for the suggestion, we have added this.

L304: Does the flatfile contain information about if the noise signal is prior to the event signal or after? In not, that’s OK, but going forward you might consider including this information in the flatfile.

- Yes it does, it is the column called ‘*noiseStart*’, this is explained in the data description file which now will be included as a supplement

L323: If possible, can you elaborate as to why ~10% of the K-NET data have “little to no usable frequency range”.

- The high number of K-NET records with low usable frequency range is likely related to low quality noise windows and fining good noise windows. This is especially difficult when the record is short, which is more often the case for K-NET (See also Figure 7). We have added a sentence about this in the manuscript.

L327: Consider adding what percentage of the data 537,550 represents.

- Thank you for the suggestion, we have added this.

L333: You mention low frequency noise a few times – any idea what might be causing this?

- The low frequency noise might be caused by several factors, for example from instrument, weather and ocean. We have added a sentence about this.

L370-371: Consider adding a sentence as to why you select a hypo-distance cut-off of 1200 km, whereas the other study goes out to 2,000.

- We use this cutoff to keep the database regional; Japanese, and avoid events that mainly occurred in other regions; e.g. the July 2025 Kamchatka event which was recorded by KiK-net and Knet stations (~1500 distance).

L408: OLD: can only be processed automatically

NEW: can *realistically* only be processed automatically

- Thank you for the suggestion, we have reformulated the sentence.

L423: Is the 'energy ratio' a new thing that you derived? Is so, I think you might want to mention that here.

- Thank you for the suggestion, we added a sentence stating this.

L430: Consider adding a percentage here.

- Thank you for the suggestion, we have added this.

L437: Someplace, maybe in the supplement, it would be nice to have a table of the values in the flatfile and a description of what they are, and an additional column listing what might constitute bad data that should be avoided.

- Thank you for the suggestion, we have added the data description file to the supplement

L591: Add a comma before Fabrice's name

- Thank you for pointing this out, we have made this correction.

Figure 1: You might consider adding something to the caption that explains why there is a peak in the data at ~200 m in Figure 1e. What causes that spike?

- Thank you for the suggestion, most KiK-net borehole are either 100 m or 200 m deep, this is related to the design of the network. We have added a sentence explaining this.

Figure 2: Nice figure! Very impressive that the deep normal faulting events clearly demark the outer-rise normal fault system.

- Thank you

Figure 3: I expect the journal will ask you to add (a), (b), (c) …. to the sub-figures.

- Thank you for the suggestion this has been added

Figure 4: Do you also include a DEMEAN step after the DETREND step? You might consider adding additional information to this figure caption. How do you compute the 'smoothed spectra'? Is the SNR computed using data from a

particular frequency band (ah, I now see you have that information in section 3.2, perhaps mention that in the caption)? How is oversampling done? A cubic spline or something else?

- Thank you for the suggestion, we have added the sentence “Each processing step is described in more detail in the Processing steps section.” to the caption.

Figure 5: Nice! If easy to do you might add a patch of yellow to the Figure 5a that is drawn on the figure first with opacity that includes the full P- and S-window duration that is used in the computations. If you do this add that information to the figure caption.

- Thank you for the suggestion, this has been added

Figure 6: Nice! You might consider adding a general sentence about how the smoothing is done.

- Thank you for the suggestion we have added this.

Figure 7: In Figure 7e – why is there a ‘Sholder’ in the KiK-Net data at around 80-100 s, that is missing in the K-net data?

- This is related to the length of the raw signal. Most KiK-net records are cut after 120 seconds. The estimated length of the signal can therefore not surpass this original length of the raw signal. Because most raw K-NET records are 60 seconds long, this “shoulder” is not seen.

Figure 8: Consider adding to this plot, when appropriate, what the expected quality thresholds are?

- Thank you for the suggestion, this has been added.

Figure 9: It might help the reader if you add a couple of sentences to this figure caption explaining the main take-home points of these figures, something similar to the last sentence in the figure caption, which I found very helpful.

- Thank you for the suggestion, we have added a sentence to the caption as well as the end of the paragraph describing the cause of the differences in cutoff frequencies.

Figure 10: Consider including N=XX to the legend for each of the different values. I expect that might help the reader, as otherwise the overprinting of the points makes it hard to decipher.

- Thank you for the suggestion, to avoid including to many repetitive values in the figure we have included the number of common values for each comparison (each row) in the caption instead

Figure S1: You might add why the comparison with the JMA catalog only goes through March, 2023?

- Thank you for the suggestion, we have added this to the caption.

Figure S3b: Why is there not data for each year? And why is there a large spike in these data?

- The distribution follows the same as the available data (Figure S1). The spike in rejected data In 2011 is related to the 11th of March Tohoku earthquake and it's aftershock sequence.

Table 1. Add more explanation as to what BA07 is referring to please.

- Thank you for the suggestion, the caption has been corrected.

Table 2. Consider adding additional information to this caption. What is the purpose and main goals of these filter parameters? What do the c, 0 and 1 subscripts represent?

- Thank you for the suggestion, we have added extra information in the caption.

List of changes from the first version of the flatfile

1. Re-download of erroneous data
    - Since the initial download (Sep., 2020) of the majority of the ground motions used to create the previous version of the dataset, NIED has announced a few corrections related to the original records. These corrections are detailed at the NIED webpage: https://www.kyoshin.bosai.go.jp/cgi-bin/kyoshin/news/newnewsindex.cgi?E. The corrections that are relevant to the first version of this dataset are mainly related to erroneous station-location of K-NET stations (Feb.01,2022, Aug.30,2022, Oct.22,2020), the withdrawn of Kik-net records (Nov.04,2025) and erroneous scale factor for KiK-net (Sep.21,2021). Due to these corrections, the related ground-motions have been re-downloaded.
2. Added records and hypocenter data
    - Ground motions and Fnet information up until end of December 2025
    - The JMA catalog is now available up until end of 2023
3. Correction of algorithm for classifying event focal mechanisms
    - There was a typo in the algorithm that misclassified events as None. Additionally, in Figure 2 in the original manuscript, only events included in the F-net catalog was included. In the new figure events without F-net match are also included (defined as None).
4. Distance calculation

- R_{hypo} calculation has been corrected.
- R_{JB} and R_{rup} is now calculated using NIED published rupture models when avaliable
- R_{epi} is now used as distance criteria instead of R_{hypo}; only records with R_{epi}<1200km are processed

5. First arrival picking algorithm
   - Now follows exactly the original method of Akazawa (2004), including step 5, which was previously omitted.
6. Corner frequency filter cutoff limits
   - Changed $f_{c0,max}$ from 0.5 Hz to 0.7 Hz and $f_{c1,min}$ from 20Hz to 15Hz
7. Computation of ROTD0 and ROTD100 PGA, PGV and PGD and response spectra (the first version only included ROTD50)
8. Some site parameters are now included in the flatfile
9. Pre-event noise is no longer used for records that start at S-wave arrival Change of borehole column names
   - All columns related to borehole recordings now ends with "_*B*", this change particularly applies to 'AriasIntensity'