Reviewer 1

Review of the paper "A Flat File Compilation of Strong Ground-Motion Intensity Measures for Crustal Earthquakes in the Indian Region"

The manuscript addresses a relevant topic within the scope of the journal, presenting a newly compiled and uniformly processed flat-file of ground-motion intensity measures (IMs) for the Indian region. This initiative represents a valuable contribution toward improving ground-motion modeling (GMM) and seismic hazard assessment in India. The manuscript has significant potential and addresses an important gap in the Indian context, however substantial improvements are needed in terms of methodological transparency and coherency, metadata completeness, and structure of the manuscript.

Major Comments

1. The decision to include records up to 600 km from the epicenter should be thoroughly justified. Most modern GMMs are validated up to about 200 km. What is the maximum fault length associated with the assumed max distance? Consider discussing the implications of anelastic attenuation and regional effects that arise including such large distances from the sources, which may bias distance scaling relationships in potential application to GMMs calibrations.

   We sincerely appreciate the reviewer's insightful comment regarding the inclusion of records up to 600 km from the epicenter. Our intent in this paper is to compile a robust, unified dataset that can serve as a foundational resource for future modeling efforts. While we acknowledge that most modern GMMs are validated within 200–300 km, including larger distances allows for broader data availability, especially in regions where strong-motion records are sparse. The fault dimensions (e.g., rupture length or width) for the distant events are not known, and therefore were not incorporated into the flat-file. Due to the lack of detailed source information for many events, distance metrics are based solely on hypocentral or epicentral distances. Nonetheless, metadata such as magnitude, epicentral location, and site characteristics were carefully curated for each record. We also acknowledge that incorporating long-distance records may introduce complexities related to anelastic attenuation and regional path effects. However, we do not analyze such effects in the present study. Including such long-distance records will be decided by GMM developers. We do want to emphasize that including such records from long distance events may still be useful in site response studies such as HVSR analysis. We have updated the manuscript with this information within lines 277 - 285 "*To ensure the most comprehensive compilation of strong-motion data for the Indian region, we have included ground-motion records up to 600 km. This facilitates the inclusion of the* *maximum number of available records, which is particularly valuable in data-scarce regions such as India. While most modern GMMs are typically validated up to 200–300 km, records at larger distances may provide useful constraints for distance-scaling behavior and attenuation trends in future regional modeling efforts. Moreover, a larger number of records may provide better constraints on site response studies such as HVSR analysis. This approach is consistent with previous efforts like the NGA-West2 flatfile, where extended-distance records were also included to improve dataset completeness.*"

2. The flat-file lacks essential finite-fault distance metrics (e.g., Rrup, Rjb, Rx), which are standard in modern GMM calibration. Given the presence of events with Mw > 5.5, it is strongly recommended to estimate such metrics using available source models or at least using empirical relationships (e.g., as in NESS2, Sgobba et al., 2021).

   We thank the reviewer for highlighting the importance of incorporating additional distance metrics using available source models or empirical relationships. As noted in Sgobba et al. (2021), the inclusion of fault geometry information in the NESS2 database either from detailed source models or virtual faults enabled the computation of various source-to-site distance metrics that are commonly required for GMM development. Unfortunately, for the Indian context, such detailed and consistent fault information is currently limited or unavailable for most events, making it extremely challenging to compute finite-fault distance measures with confidence. Moreover, we do not have reliable information on which event is definitively associated with which fault. There is considerable uncertainty in these associations, and this lack of clarity further limits the applicability of source-based distance metrics. Even empirical approaches, such as those proposed by Kayastha et al. (2023), require input parameters like fault azimuth and rupture dimensions data that are typically not reported or are highly uncertain for the majority of Indian earthquakes. As a result, until such information becomes more readily available or systematically compiled, it is not feasible to include these additional distance measures in the current version of the dataset.

   We have incorporated this within the lines 291 - 301 "*It is important to note that detailed fault rupture information such as fault geometry, rupture dimensions, and slip distribution is currently limited or unavailable for the majority of the events in our dataset. This limitation poses a significant challenge in accurately computing finite-fault distance measures (e.g., rupture distance or Joyner-Boore distance), which require well-constrained rupture parameters. As a result, our flat-file includes only epicentral and hypocentral distances. While these measures may not fully capture the source-to-site geometry for large-magnitude events, they remain the most consistent and widely available metrics across the dataset, ensuring a uniform and practical basis for ground motion analysis in the absence of detailed rupture modeling. Moreover, we do not have* *reliable information on which event is associated with which fault. There is considerable uncertainty in these associations, and this lack of clarity further limits the applicability of source-based distance metrics".*

3. All the original magnitude types should be included in the flat-file. Indeed, the original estimates should be preserved for transparency. The homogenization to Mw should be accompanied by clear documentation of the applied conversion equations and associated uncertainty.

We fully agree that retaining the original magnitude types reported by various agencies is essential for transparency and reproducibility. In the flat-file, this information has been preserved in the column labeled "Given_Mag," with the IGM_flat-file which documents the original magnitude values along with their corresponding types as provided by the different agencies. Furthermore, the process of homogenizing magnitudes to moment magnitude (Mw) has been clearly described in Part I of the manuscript. The empirical relationships used for conversion are cited and documented in lines 204-205, ensuring clarity regarding the applied methodology. Also, we have now updated the manuscript line 195 "*each scale is mentioned in IGM flat-file"*. While we acknowledge that magnitude conversions inherently involve uncertainties, especially when applied across a diverse dataset, the chosen empirical relationships were selected based on their relevance to regional data and previously validated performance.

4. Although Vs30 is included in the station metadata, its derivation and role in the flatfile structure are not discussed consistently across Sections 3.1 and 3.2. This should be clarified. Moreover, relying solely on generalized surface geology or assumed values (e.g., 760 m/s for rock) is insufficient. Whenever possible, values should be derived from geophysical investigations or well-documented proxies.

We sincerely thank the reviewer for the insightful comment regarding the treatment of V_{S30} in the flat-file. We are in full agreement with the reviewer that the site response parameters such as V_{S30} should be derived based on geophysical measurements however this is beyond the scope of the current study. Ideally such information should be provided by the agencies which maintain and operate the various seismic networks within the country. Moreover, if such measurements are made for each station site the related geophysics data (e.g. Vp and Vs profiles) should be made available. However, we do hope such practices will evolve in the near future and this study is an attempt in this direction. In response, we have clarified this aspect in the manuscript, specifically in Lines 159-165 *"For site information particularly V_{S30} the values provided in the flat-file are directly extracted from the available metadata accompanying the original strong-motion records, as reported by the respective networks. In most cases, the* *networks did not document the methodology used to derive these V_{S30} values, whether through direct geophysical measurements, proxy-based estimations, or generalized geological assumptions. Consequently, this information could not be further elaborated in the flatfile. To enhance the site metadata, surface geology classifications were additionally compiled from relevant literature, with the aim of including as much site-specific information as possible."* Furthermore, we have clearly outlined the available site condition details, including classifications such as rock or soil types, in Part I of the manuscript under the description of each dataset. We acknowledge the limitations associated with assumed or unverified V_{S30} values and highlight the need for future efforts to acquire more rigorously derived site parameters through direct investigations or validated proxy methods.

5. The authors state that the dataset spans multiple tectonic regimes but do not define or distinguish them in the flat-file. This is a critical omission, as GMMs are tectonic regime specific. A field indicating tectonic classification is essential.

   We thank the reviewer for this important observation. In response, we have revised the manuscript throughout to clearly specify the tectonic setting relevant to our study. While the dataset was initially described as covering a broader region, we have now consistently clarified that the focus is specifically on the Himalayan tectonic regime. To avoid any ambiguity, we have updated Figure 1 to better reflect the geographic and tectonic extent of the dataset, ensuring alignment between the described study region and the data presented. We agree that including a field indicating tectonic classification is valuable. In the current version, as the dataset is limited to the Himalayan region, a single regime is applicable. However, future extensions of this dataset may include multi-regime classification as more data becomes available from other regions in India. Lines 73 - 77 have been updated in the manuscript "*Given this disparity, the primary motivation of this study is to develop a uniformly processed ground-motion flat-file, tailored to the Indian context, with a particular focus on the Himalayan region. The decision to focus on this area stems from the comparatively higher data availability, which enables more robust analysis and model development"*

6. The manuscript does not discuss whether multiple recordings from the same station were analyzed for potential site-specific biases or dependence and for potential implementation of non-ergodic approaches. Since several stations (for example from the PESMOS network) appear to have recorded many events, the flat-file may include the number of records per station.

   We thank the reviewer for making this important scientific point. As rightly noted, several stations particularly those from the PESMOS network have recorded multiple events, which could introduce site-specific effects and dependencies relevant for non-ergodic ground motion modeling. In response, an additional sheet has been added to the flat-file that lists each unique station name along with the corresponding number of records, thereby providing a clear summary of data availability per station. Also, each station has been assigned a unique identifier in the main flat-file, incorporating network-specific station codes. This allows users to easily track and analyze all recordings from the same site across different networks. We believe these additions will facilitate more informed analyses of site-specific effects and support the future implementation of partially non-ergodic modeling approaches.

We would also like to clarify that this study is not focussed on any modelling aspects. For sure there will be repeatable site and path effects present in the data as rightly noted by the reviewer however, these issues come next when the data will be used for developing GMMs and they will need to be carefully accounted for. Thanks again for pointing this out and we have added a sentence in the manuscript within the line 768-770: "*Given some of the stations in our dataset (mainly from PESMOS) have recorded multiple events, in the development of GMMs the repeatable site (and path) effects will need to be accounted for where data permits*".

7. There are several inconsistencies between Sections 3.1 and 3.2. For example, key details about metadata, filtering, and IM calculation are inconsistently presented. The structure of these sections should be revised to improve coherence.

   We thank the reviewer for pointing out the inconsistencies between Sections 3.1 and 3.2. In response, both sections have been restructured and revised to ensure greater clarity, consistency, and coherence in the presentation of metadata details and the computation of intensity measures (IMs). We have carefully reviewed the content to eliminate redundancies and align the structure of these sections, thereby improving the overall readability and flow of the manuscript. We appreciate the reviewer's observation, which helped us significantly enhance the quality of this part of the paper.

8. The methodology for defining low- and high-cut frequencies based on SNR is described in general terms but lacks clarity on whether fixed thresholds, adaptive criteria, or visual inspection were used. This ambiguity could affect the reproducibility and reliability of the processed IMs, especially for long-period estimates.

   We thank the reviewer for raising this important point regarding the methodology used to define the low- and high-cut filter frequencies based on signal-to-noise ratio (SNR). To clarify, the process implemented in our study is fully automated and based on an adaptive thresholding approach that ensures both objectivity and reproducibility. No fixed cut-off frequencies or manual visual inspections were used at any stage of filtering. Instead, the lower and upper corner frequencies (fc_low and fc_up) were determined individually for each waveform using a consistent algorithm based on the frequency-dependent computation SNR of the signal at each frequency ordinate. We have made the necessary changes and added within the manuscript, specifically in lines 479-516 *"The noise and signal windows were identified using a P-phase picker algorithm based on the Akaike Information Criterion (AIC), following the approach of Kalkan (2016). This method enabled the automated and consistent detection of seismic phases namely, the noise window, P-wave onset, and S-wave arrival for each waveform. The noise window was defined as the portion of the waveform extending from the beginning of the record up to the P-wave arrival, typically corresponding to the first significant amplitude increase. The signal window was taken as the segment following the P-wave arrival, encompassing the S-wave energy. To prepare the signals for frequency-domain analysis, both windows were subjected to zero-padding to avoid spectral leakage and improve frequency resolution. The resulting amplitude spectra were then smoothed using the Konno and Ohmachi (1998) logarithmic smoothing algorithm with a bandwidth parameter of b = 40, which ensures stability and preserves the key spectral features of the signal and noise.*

The signal-to-noise ratio (SNR) is calculated by dividing the smoothed root mean square (RMS) amplitude of the signal by the smoothed RMS amplitude of the noise within specific frequency bands. This calculation is performed iteratively over progressively adjusted frequency windows until the SNR exceeds a threshold value of 3. For each frequency window, the mean squared amplitudes of the signal and noise are computed, and their square roots are taken to obtain the respective RMS values. The SNR is then determined as the ratio of the signal RMS to the noise RMS. This procedure is applied independently at both the low and high ends of the frequency spectrum to determine the appropriate cutoff frequencies (fc_low and fc_up). To identify the low-cut frequency, the algorithm begins at a predefined lower bound (e.g., 0.1 Hz) and increases the frequency incrementally (e.g., in 0.1 Hz steps), evaluating the SNR at each step. The lowest frequency at which the SNR surpasses the threshold of 3 is selected as the low-cut frequency (fc_low). Similarly, the high-cut frequency is determined by starting from a higher frequency bound (e.g., 45 Hz) and stepping downward until the SNR again exceeds the threshold, with that frequency recorded as the high-cut frequency (fc_up). This adaptive and automated approach ensures that the filter band is specifically tailored to the signal quality of each individual waveform, allowing for the retention of reliable frequency content while excluding portions dominated by noise. The method avoids reliance on arbitrary fixed bands or visual inspection, thereby improving the objectivity, reproducibility, and robustness of the processing. In particular, by basing the cutoff frequencies on the actual spectral characteristics of the data, this technique enhances the reliability of IM calculations especially at long periods, where low-frequency noise can otherwise introduce significant bias. Inadequate low-frequency filtering may suppress long-period content and underestimate IMs at longer periods, while overestimation of *high-frequency noise could bias PGA or short-period SA. By defining the frequency limits based on actual signal conditions rather than fixed assumptions, the method ensures that IMs are computed from reliable and physically meaningful portions of the waveform, enhancing both accuracy and scientific credibility"*.

9. While the manuscript refers to standard processing protocols (Puglia et al., 2018), it remains unclear which specific software packages or processing scripts were used (ObsPy, custom scripts?). A clear description of the toolchain, including parameter settings, would enhance transparency and reproducibility.

   We appreciate the reviewer's suggestion to clarify the software tools and processing environment used in this study. We have added these lines 424 - 430 "*The complete list of software packages, data repositories, and processing scripts has been provided in the Data and Code Availability section of the manuscript. Specifically, the waveform processing and ground-motion parameter extraction were carried out using standard libraries including ObsPy (Beyreuther et al., 2010), NumPy, and Pandas, while figures were generated using PyGMT (Tian et al., 2024). The processing workflow follows the signal processing protocols outlined in Puglia et al. (2018), implemented through custom Python scripts built on these open-source libraries"*.

10. The manuscript lacks discussion on how the filtering process may affect long-period spectral displacement estimates. This is particularly relevant for periods >5s, where baseline correction and residual drift removal may significantly bias results.

    We appreciate the reviewer's observation regarding the potential impact of filtering on long-period spectral displacement estimates, particularly for periods greater than 5 seconds. This clarification has been incorporated into the manuscript and is now explicitly addressed in lines 533 - 548 "*In our study, we applied causal Butterworth bandpass filters to the acceleration time histories, following standard protocols as described in Boore and Akkar (2003) and Boore and Bommer (2005), with careful consideration to avoid phase distortion and preserve the true signal characteristics. The low-cut filter was specifically designed to remove low-frequency noise that can adversely affect long-period spectral components, while the high-cut filter was set based on the Nyquist frequency, which depends on the sampling rate of each record. To minimize the inclusion of unreliable long-period content, we excluded waveforms with coarse sampling intervals (i.e., $\Delta t > 0.05$ s), recognizing that such sensors often have limited bandwidth and may inaccurately capture peak motion or long-period trends (Douglas and Boore, 2011). Additionally, a strict SNR threshold of 3 was used to determine filter corner frequencies, ensuring that only signals with sufficient energy above noise were retained. We acknowledge that long-period ground motion is particularly sensitive to issues such* *as baseline shifts and residual drift, which can bias displacement time series and spectral displacements. Although explicit baseline correction procedures were not applied in this version of the dataset, the filtering strategy and waveform selection criteria were tailored to mitigate such effects indirectly by removing noisy or poorly resolved low-frequency signals*". To promote transparency, we have annotated the flat-file with key metadata including the sampling interval, low- and high-cut frequencies, allowing users to assess the suitability of each record for long-period analysis. As per the reviewer's suggestion, we have also plotted the Fourier Amplitude Spectrum (FAS) for both the raw (unfiltered) and filtered acceleration time histories (see Figure below) for one of the records from PESMOS. The same was carried out for all the records. The unfiltered signal (red curve) exhibits elevated amplitudes at low frequencies (< 0.2 Hz), which are characteristic of long-period noise and baseline drift. After applying the causal Butterworth bandpass filter, the filtered signal (blue curve) shows a clear attenuation of this low-frequency noise. Notably, the filtered FAS demonstrates the expected decay pattern consistent with the ω^2 (f^{-2}) shape of earthquake source spectra at higher frequencies, confirming the physical plausibility of the retained signal. This plot supports our filtering strategy, demonstrating that the applied bandpass filter effectively suppresses spurious low-frequency content while preserving the true spectral characteristics of the ground motion. A similar procedure was followed for all the records, and the corresponding results were analysed.

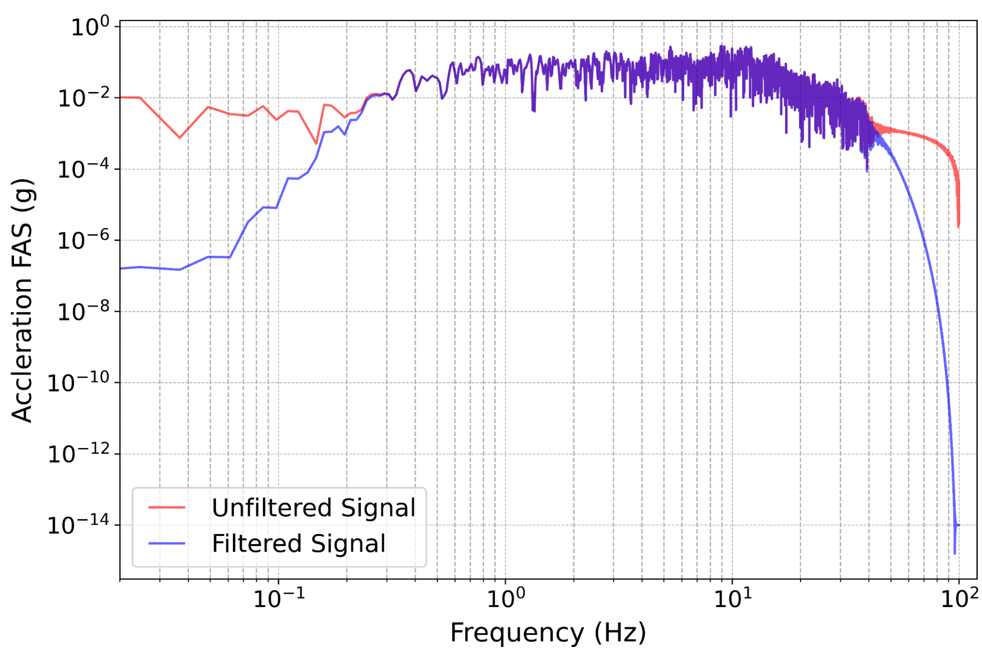

Figure. Comparison of the Fourier Amplitude Spectrum (FAS) for unfiltered (red) and filtered (blue) acceleration time histories. The unfiltered signal shows elevated low-frequency content due to noise and baseline trends.

11. The manuscript does not include any statistical analysis or visual summary of the IM distributions. Histograms, boxplots, or scatter plots of PGA, PGV, SA, and duration metrics would enhance the understanding of the database's representativeness.

    We sincerely thank the reviewer for this valuable suggestion. In response, we have included a set of visual summaries scatter plots illustrating the variation of PGA and SA (at T = 0.5 s and 1.0 s) with magnitude and distance to provide better insight into the distribution and representativeness of the intensity measures in the compiled database. These figures have been added to the revised manuscript and are now discussed in detail in lines 326-334 "*Figure 4 presents the distribution IMs as a function of magnitude and hypocentral distance for the compiled dataset. Figure 4(a) shows variation of PGA with moment magnitude (Mw), while figure 4(b) illustrates the expected attenuation of PGA with increasing hypocentral distance (RHypo). Figure 4(c) and (d) display the variation of SA at 0.5 s and 1.0 s, respectively, as a function of distance. In both cases, SA shows a similar decay trend with distance, although a larger scatter is observed at longer periods due to the increasing influence of site conditions and signal-to-noise issues. These plots highlight the distribution and variability of IMs across different source characteristics and propagation distances, underscoring the need for robust ground motion prediction models*".

12. The validation using Bora et al.'s model is inadequate. What are the validity limits (magnitude, distance, tectonic setting) of this model? Are they compatible with the dataset? Further, decomposition of residuals into between- and within-event terms is essential to identify systematic biases with magnitude and distances.

    We thank the reviewer for this insightful comment. The Bora et al. (2019) model, used in our sensitivity analysis, is one of the only two internationally developed empirical models for Fourier Amplitude Spectrum (FAS), the other being Bayless and Abrahamson (2019). We selected the Bora et al. model primarily due to its relatively simpler functional form and ease of implementation, which makes it more suitable for initial sensitivity checks and comparison with Indian recordings, especially given the current absence of well-established region-specific FAS models. The Bora et al. model was developed using a subset of the NGA-West2 database, primarily consisting of recordings from active tectonic regions such as California. It is valid for moment magnitudes ranging from approximately 3.0 to 7.5, source-to-site distances up to 300 km, and a frequency range of

0.1 to 45 Hz. These validity limits are broadly consistent with our compiled dataset. Moreover, the dataset utilized in this manuscript primarily represents the Himalayan region, which is also an active tectonic setting, making the application of the Bora et al. model contextually reasonable for preliminary analysis.

We used the Bora et al. model also for the preliminary validation because it reflects common practice in Indian ground-motion studies, where imported models are often employed in the absence of indigenous alternatives. The sensitivity analysis results (discussed in Sensitivity analysis and shown in Figure 11a and 11b) highlight systematic residual trends, particularly between 1–10 Hz, where observed Indian motions are generally lower than model predictions. We fully agree with the reviewer that decomposing residuals into between-event and within-event components is essential for diagnosing potential biases with respect to source, path, or site effects. However, as this manuscript is focused on presenting the flatfile and its structure, we chose not to delve into full GMM evaluation or calibration here. A detailed decomposition analysis, including mixed-effects regression and regionally adjusted GMM development for the Indian context, has been carried out in a separate study.

Minor and Editorial Comments

Line 50: Reference for the Italian flat-file should cite Lanzano et al., 2019 (ESM database).

We thank the reviewer for pointing out the missing reference. We have now added the citation to the Engineering Strong Motion (ESM) database by Lanzano et al. (2019), along with other relevant studies. The updated manuscript includes the following revised line 50-56 *"Additionally, individual studies also targeted development of such flat-files for ground-motion IMs that involve Akkar et al. (2010) for Turkey, Arango et al. (2011a) for the Central American subduction zone, Pacor et al. (2011) for the ITACA database in Italy, and more recently, Lanzano et al. (2019) developed the Engineering Strong Motion (ESM) flat-file for Europe and the Middle East, providing a harmonized dataset with uniformly processed recordings and metadata critical for pan-European seismic hazard assessments"*. We appreciate the reviewer's suggestion, which helped improve the completeness and accuracy of our references.

Line 76: The acronym "GMC" is introduced but not used thereafter. please revise or remove.

We thank the reviewer for noting this inconsistency. The acronym "GMC" was not used subsequently in the manuscript, and we have therefore removed it to improve clarity and consistency.

Lines 113–119: Referring to manuscript sections as "Part I", "Part II", etc., is unnecessary and can be confusing.

We thank the reviewer for this helpful suggestion. To avoid any confusion and improve clarity for the readers, we have replaced the references to "Part I", "Part II", "Part III" and "Part IV"., with the full descriptive titles of the respective studies. Specifically, the manuscript now clearly refers this within the lines 166 - 172 **"***Part I INDIAN GROUND MOTION (IGM) DATABASE", Part II IGM PROCESSING PROTOCOL, Part III METADATA, Part IV ADDITIONAL GROUND-MOTION INTENSITY MEASURES (IMs)***"**. This revision ensures that the purpose and scope of each study are clearly communicated to the reader without relying on shorthand labels.

Lines 144–146: Avoid replacing all magnitudes with Mw estimates. Include both original and converted values in separate columns, as discussed in the general comments

We thank the reviewer for thoughtfully highlighting this important point. As suggested, we have addressed this in both the general comments and the manuscript. Specifically, we now include both the original reported magnitudes and their converted moment magnitude (Mw) estimates in separate columns within the flat-file. The actual magnitude scale used for each record is also clearly indicated, based on the reporting conventions of the respective networks. This approach preserves the integrity of the original data while also providing consistency for comparative analysis.

Line 243–245: Fixed time windows may not capture signal appropriately across varying magnitudes/distances. A duration model would be more robust.

We sincerely thank the reviewer for this important observation. We fully agree that using a duration model provides a more robust and physically informed approach to signal windowing, particularly across a range of magnitudes and distances.

In this preliminary stage, our primary aim was to compile and standardize a comprehensive strong-motion dataset for the Indian region. Therefore, we initially applied fixed time windows to ensure consistent signal processing and to minimize the inclusion of pre-event and post-event noise. However, we acknowledge the limitations of fixed windows, especially in relation to varying event characteristics. To further explore this, we conducted a supplementary analysis of the expected significant duration of each record using Husid plots. These plots helped assess whether the applied windows were adequate in capturing the full duration of strong motion. We recognize that implementing a formal duration model (e.g., Afshari & Stewart, 2016) would offer a more systematic method. However, such models require additional distance metrics such as rupture distance (Rrup) which are currently unavailable for many records in our dataset due to incomplete source characterization. As such, while duration-based windowing is highly relevant, its full implementation will be a focus of future work, particularly in studies aimed at ground-motion model calibration and duration scaling.

Lines 247–248: Specify filtering details—type, order, cutoff frequencies. Were records visually inspected? Was tapering applied? This is discussed later but mentioned in a rough way here.

We thank the reviewer for this valuable comment. To improve clarity and transparency, we have revised the section on waveform filtering to provide more specific details on the processing steps. In particular, we have now clearly described the filter type, order, cutoff frequencies, visual inspection of records, tapering, and other preprocessing steps. The revised text (lines 344–351) now reads "*All waveforms were visually inspected to ensure the presence of clear P- and S-wave arrivals. A Butterworth bandpass filter was applied to remove high and low-frequency noise, minimizing phase distortion (Boore and Akkar 2003; Boore and Bommer 2005). The high-cut frequency was set near the Nyquist limit, while the low-cut filter removed long-period noise. Records with sampling intervals < 0.05 s were excluded to avoid bandwidth-related issues (Douglas and Boore 2011). A cosine taper (5% of total duration) and zero-padding were applied at both ends of the records (Converse and Brady 1992). Instrument response correction and baseline trend removal were also performed*."

Vertical Component Exclusion: Consider including vertical ground motions, which are increasingly relevant for infrastructure modeling.

We thank the reviewer for this valuable suggestion regarding the inclusion of vertical ground motions. In response, we have now included a dedicated file named IGM_Flat-file_Vertical, which contains the vertical-component ground-motion records with all relevant intensity measures computed using the same processing methodology as for the horizontal components.

We have also updated the manuscript to clarify this inclusion. The lines 355 - 359 text now states *Vertical-component ground motions are also included in our present analysis although the vertical-component ground motions were not used in the computation of scalar or rotated intensity measures (in the present study), they were processed using the same methodology and are included in the flat-file as a separate worksheet (IGM Flat-file_Vertical) for completeness and future use*."

Line 258: Why was only NCS data "extensively processed"? This comment lacks context—please clarify.

We thank the reviewer for pointing out the need for clarification. The statement has now been revised in the manuscript to better explain why only the NCS data were described as "extensively processed." Specifically, the records from the National Center for Seismology (NCS) were provided in raw form, without any prior corrections or preprocessing. As such, we carried out the complete processing workflow manually including baseline correction, instrument response removal, filtering, and careful visual inspection of each waveform. We also have revised the manuscript lines Lines 368 - 372 “*because the records were provided in raw form, without any prior corrections. Therefore, we carried out the entire preprocessing manually; this includes baseline correction, filtering, instrument response removal, and visual inspection of each waveform. However, many events were excluded due to unclear identification of P and S-wave arrivals in the waveforms”*.

Line 265: Describe the method for picking P and S phases.

We thank the reviewer for this helpful comment. In response, we have now updated the manuscript to include a description of the method used for picking P- and S-wave arrivals. The revised text has been added in the reviewed manuscript.

Lines 397-403 “*P- and S-phase arrivals were identified using a P-phase time picker based on the Akaike Information Criterion (AIC) method, as implemented in the MATLAB code developed by Kalkan (2016). This algorithm detects the onset of seismic phases by minimizing the AIC function, which distinguishes between noise and signal by evaluating changes in the statistical properties of the waveform. The method was applied to each acceleration time history to accurately determine the signal window, and the picks were visually verified to ensure consistency and reliability”*.

Lines 479 - 490 “*The noise and signal windows were identified using a P-phase picker algorithm based on the Akaike Information Criterion (AIC), following the approach of Kalkan (2016). This method enabled the automated and consistent detection of seismic phases namely, the noise window, P-wave onset, and S-wave arrival for each waveform. The noise window was defined as the portion of the waveform extending from the beginning of the record up to the P-wave arrival, typically corresponding to the first significant amplitude increase. The signal window was taken as the segment following the P-wave arrival, encompassing the S-wave energy. To prepare the signals for frequency-domain analysis, both windows were subjected to zero-padding to avoid spectral leakage and improve frequency resolution. The resulting amplitude spectra were then smoothed using the Konno and Ohmachi (1998) logarithmic smoothing algorithm with a bandwidth parameter of b = 40, which ensures stability and preserves the key spectral features of the signal and noise*”.

Lines 415–417, 444: Clarify whether rotated IMs include vertical components. Why is RotD0 excluded?

We thank the reviewer for this valuable comment. We would like to clarify that all intensity measures (IMs), including rotation-independent parameters and duration measures, were computed using only the two horizontal components. Vertical components were not included in the computation of rotated IMs. However, the vertical component records have now been processed using the same methodology and are included in the flat-file as a separate worksheet for completeness and future use. Additionally, we acknowledge the omission of RotD0 in the earlier version and have now included it in our analysis. The manuscript has been updated accordingly to clearly reflect these additions and distinctions

Lines 427–447: Definitions for duration measures (D5–75, etc.) and EAF appear inconsistently. Introduce references properly when first mentioned.

We thank the reviewer for this helpful observation. In response, we have revised the manuscript to ensure that the definitions of duration measures (e.g., $D_{5–75}$, $D_{20–80,}$, $D_{5–90}$) and the Effective Amplitude Function (EAF) are introduced more clearly and consistently. Relevant references have now been properly cited at the point of first mention to provide context and improve clarity for the reader.

These changes have been made in lines below, and we believe they enhance the readability and technical accuracy of this section. We appreciate the reviewer's suggestion, which helped us strengthen the overall presentation.

Lines 680- 694 "*Ground-motion durations were quantified using significant duration metrics based on the accumulation of AI, following established practices in engineering seismology (Trifunac and Brady 1975; Kempton and Stewart 2006). These metrics are widely used to represent the time intervals over which the majority of the seismic energy is delivered, which has implications for structural response and damage potential.*

$D_{5–75}$: *The time interval between the points when 5% and 75% of the total AI is reached. This represents the time span over which the central 70% of the energy is released.*

$D_{20–80}$: *The time interval between 20% and 80% of AI accumulation. It captures a narrower energy window, useful in studies focusing on the peak energy release.*

$D_{5–95}$: *The time interval between 5% and 95% of AI, encompassing nearly the entire duration of significant shaking and often used in dynamic analyses of structures.*

These duration measures are calculated from the cumulative Arias Intensity curve and provide robust estimates of ground-motion duration that are independent of amplitude thresholds, making them particularly useful for comparing motions across a wide range of magnitudes and distances (e.g., Bommer et al., 2009; Afshari and Stewart 2016)".

Lines 698 - 705 "*The EAS represents the amplitude-modulated frequency content of a ground-motion record and is computed as the geometric mean of the FAS of two horizontal components, smoothed to reduce high-frequency variability (Boore, 2003). In this study, EAS was calculated over a frequency range of 0 to 40 Hz to comprehensively capture the frequency characteristics relevant for engineering seismology and seismic hazard analysis. Unlike raw FAS, the EAS provides a more stable and representative measure of ground-motion frequency* *content by applying a smoothing window (e.g., Konno–Ohmachi), making it particularly useful for site response analysis, simulation validation, and spectral shape characterization (Boore, 2005; Atik et al., 2014*)".

Lines 484–489: Clarify which processing steps were followed—Puglia et al. (2018) or otherwise? Reconcile with earlier processing descriptions.

We thank the reviewer for this important comment. To address the inconsistency and improve clarity, we have revised the manuscript to clearly specify the processing steps followed and to ensure consistency throughout the text. Specifically, we now state that the ground-motion recordings were processed following the standardized protocols outlined by Puglia et al. (2018) and Luzi et al. (2016), as detailed in Part II of the study. The revised text now explicitly describes each step of the processing workflow including baseline correction, second-order acausal Butterworth filtering, record-specific cutoff frequencies (documented in the flat-file), tapering, zero-padding, and spectral smoothing using the Konno and Ohmachi (199**8)** algorithm. We have also clarified that visual inspection and manual quality control were performed in accordance with Luzi et al. (2016). These revisions ensure a consistent and transparent description of the processing methodology across the manuscript. We appreciate the reviewer's comment, which has helped improve the overall clarity and coherence of our submission.

Lines 784 - 796 *"The ground-motion recordings were manually processed following the standardized protocols outlined by Puglia et al. (2018) and Luzi et al. (2016), as detailed in Part II. Processing began with baseline correction, followed by the application of a second-order acausal Butterworth filter, using carefully selected low- and high-cut frequencies specific to each record. These cutoff frequencies are documented in the flat-file accompanying the dataset. To ensure numerical stability and consistency between acceleration and displacement after double integration, tapering and zero-padding were applied at the beginning and end of each acceleration time series. Additionally, spectral smoothing was performed using the Konno and Ohmachi (1998) algorithm to stabilize the frequency-domain representations such as the EAS. Each record underwent visual inspection and manual quality control, consistent with the procedures described in Luzi et al. (2016), to exclude data affected by clipping, transient noise, or poor signal-to-noise ratio. This rigorous processing strategy ensures a high-quality, uniformly processed database suitable for seismic hazard assessment and engineering applications".*

Equation 5: Correct the lower integral limit (should start from 0).

We thank the reviewer for pointing out this error. The lower limit of the integral in Equation 5 has now been corrected to start from 0, as suggested. This correction has been made in the revised manuscript.

Reviewer 2

The effort undertaken to develop a ground-motion intensity measure (IM) flatfile for the Indian region is commendable. The compilation of a uniformly processed and high-quality dataset aligned with international standards such as NGA is a significant contribution. The authors have carefully processed and validated approximately 854 strong-motion records, incorporating a wide range of IMs that are critical for ground-motion model (GMM) development and seismic hazard assessment. This is a much-needed initiative in the Indian context and will serve as a valuable resource for researchers, practitioners, and policymakers involved in earthquake engineering and seismic risk mitigation. Given the broader significance of this work, it is essential to establish a robust and reliable foundation by thoroughly addressing the issues outlined below. Reviewer recommends a major revision.

Major Comments

1. Figure 1 shows earthquake epicenters and seismic stations across the Indian subcontinent but lacks key details for full interpretation. There is no magnitude scale to explain the varying sizes of the epicenters, nor is there information about the time frame of the data. A distance scale, elevation legend, and tectonic features like fault lines or plate boundaries are missing. The types of seismic stations are also unspecified. Additionally, the map lacks a title, region or city labels, and references to data sources. These omissions limit the map's clarity, scientific value, and usefulness for understanding regional seismicity.

   We sincerely thank the reviewer for their insightful comment regarding Figure 1. We agree that these elements are essential for enhancing the clarity and interpretability of the map. Accordingly, we have thoroughly revised Figure 1 to address all the mentioned concerns. The updated figure now includes a magnitude scale for the epicenters, a distance scale bar, a legend indicating elevation, clearly marked tectonic features such as fault lines and plate boundaries, and labels for major regions and cities. These changes are reflected in the revised manuscript updated Figure 1.

2. The limitations should be acknowledged in the manuscript. The dataset only covers records up to the year 2018, and no information is provided regarding the number of earthquakes per year, which is important for assessing temporal completeness and trends in seismicity.

   Thank you for this valuable suggestion. As suggested, we have now included a histogram (Figure 3) that displays the number of earthquake records per year in our compiled dataset with lines 309-312 "*The histogram in figure 3 illustrates the number of earthquakes recorded each year in the compiled dataset. A significant increase in the* *number of events is observed after 2008, likely due to the densification of seismic networks, enhanced instrumentation, and improved reporting protocols. The sparse data in earlier years reflect limitations in catalog completeness and station coverage*". We believe this visualization enhances the clarity of the manuscript and addresses the concern regarding temporal trends and completeness.

3. Several important parameters relevant to seismic design and performance assessment—such as Acceleration Spectrum Intensity (ASI), Velocity Spectrum Intensity (VSI), Characteristic Intensity (Ic)—are not included. Authors are encouraged to refer to recent studies, such as https://doi.org/10.1016/j.istruc.2023.06.007 and https://doi.org/10.1016/j.soildyn.2024.108923.

We thank the reviewer for pointing out the importance of incorporating additional intensity measures for comprehensive ground-motion characterization. As suggested, we have now computed and included the following intensity measures in the flat-file: Acceleration Spectrum Intensity (ASI), Velocity Spectrum Intensity (VSI), and Characteristic Intensity (Ic). These additions are now explicitly described in the revised manuscript.

Line 589 - 592 "We also derived Acceleration Spectrum Intensity (ASI) and Velocity Spectrum Intensity (VSI), which are increasingly recognized in engineering practice for their relevance in structural analysis. Alongside these, we included Characteristic Intensity (Ic), an indicator of potential structural damage".

Lines 640-643 "Furthermore, we calculated ASI and VSI, which have recently gained prominence for their frequent use in engineering applications In addition, we evaluated the Characteristic Intensity (Ic), an intensity measure associated with structural damage potential due to maximum deformation and hysteretic energy absorption".

Lines 706-721 "In addition to conventional ground-motion parameters, we computed ASI, VSI and Ic which have recently gained traction for their utility in seismic performance assessments (Sriwastav et al., 2024, Dai et al., 2023). ASI is calculated as the integral of the 5% damped pseudo-acceleration response spectrum (PSA) over a period range of 0.1 to 0.5 seconds, capturing the intensity relevant to short-period structures such as low-rise buildings (Thun et al., 1988). Similarly, VSI is obtained by integrating the pseudo-velocity response spectrum (PSV) over the period range of 0.1 to 2.5 seconds, making it useful for assessing the impact of ground motions on mid- to long-period structures (Housner, 1959). These spectral intensity measures directly reflect the engineering demand represented by the shape and area under the uniform hazard spectrum (UHS). Additionally, we computed I_c*, which combines both the amplitude and* *duration characteristics of ground motion (Liao et al., 2022). It is derived as a function of the root mean square acceleration (A_{rms}) and the significant duration (T_{sig}), and is considered an indicator of structural damage potential due to its sensitivity to maximum deformation and energy dissipation through hysteresis (Park and Ange, 1985). These advanced parameters are included in the flat-file and are intended to support a wide range of engineering applications, particularly those focused on performance-based seismic design and vulnerability assessment".*

Lines 929-952; we have included the following definitions and rationale*:*

c) Acceleration Spectrum Intensity (ASI) is computed as the integral of the pseudo-spectral acceleration (PSA) over the period range 0.1- 0.5 s (Thun et al., 1988) for dam safety evaluation. ASI has gained relevance in modern seismic engineering due to its ability to capture the energy content of ground motions affecting short-period structures. It is particularly beneficial for seismic performance-based loss estimation in urban environments dominated by low- to mid-rise buildings. ASI is computed using the following expression:

$$ASI = \int_{0.1}^{0.5} PSA(\zeta = 0.05, T)dT \qquad (7)$$

d) Velocity Spectrum Intensity (VSI) integrates the pseudo-spectral velocity (PSV) over a longer period range of 0.1–2.5 s (Housner G. W. (1959)). VSI is useful for characterizing the shaking intensity experienced by longer-period structures, including mid- to high-rise buildings and bridges. It is given by:

$$VSI = \int_{0.1}^{2.5} PSV(\zeta = 0.05, T)dT \qquad (8)$$

e) Characteristic Intensity (Ic) captures the energy content of a ground motion through a combination of root mean square acceleration (A_{rms}) and significant duration (T_{sig}) (typically between the 5% and 95% Arias intensity markers). The expression for Ic is:

$$Ic = (A_{rms})^{1.5} \cdot (T_{sig})^{0.5} \qquad (9)$$

This parameter is valuable for correlating with structural damage and energy-based metrics.

Added References

Von Thun, J. L. (1988). Earthquake ground motions for design and analysis of dams. Earthquake engineering and soil dynamics II-recent advances in ground-motion evaluation.

Housner, G. W. (1959). Behavior of structures during earthquakes. Journal of the Engineering Mechanics Division, 85(4), 109-129.
Sriwastav, R. K., Yedulla, J., & Raghukanth, S. T. G. (2024). A non-parametric model of ground motion parameters for shallow crustal earthquakes in Europe. Soil Dynamics and Earthquake Engineering, 186, 108923.
Dai, J. C., Wang, D. S., Chen, X. Y., Zhang, R., & Sun, Z. G. (2023, September). Evaluation of ground motion intensity measures for time-history dynamic analysis of isolated bridges. In Structures (Vol. 55, pp. 1306-1319).
Liao, B. Y., Huang, H. C., & Xie, S. (2022). The source characteristics of the Mw6. 4, 2016 Meinong Taiwan earthquake from teleseismic data using the hybrid homomorphic deconvolution method. Applied Sciences, 12(1), 494.
Park, Y. J., & Ang, A. H. S. (1985). Mechanistic seismic damage model for reinforced concrete. Journal of structural engineering, 111(4), 722-739.

4. The reviewer believes that each of the data are not accessed carefully by the authors. Several of the PESMOS data have suffered from clipping but are included in the manuscript. Such clipped data may demonstrate reliable response spectra and IMs. However, are not correct and should be excluded. An example is provided below.

We sincerely thank the reviewer for highlighting this important point regarding data quality, particularly in relation to the inclusion of clipped PESMOS records. We would like to clarify that the main focus of this study is on the two horizontal components (H1 and H2) of recorded ground motions. All derived ground-motion parameters such as response spectra, Fourier amplitude spectra, and intensity measures are computed using the geometric mean of these horizontal components. This approach reflects common practice in ground-motion model development and seismic hazard analysis, where horizontal motion is typically prioritized due to its primary role in structural response. However, in response to the reviewer's suggestion and to ensure completeness, we have now also included a separate flat-file for the vertical component. This will allow future users to explore vertical ground motion if required, although our primary analysis remains focused on the horizontal components.

We fully agree that clipped records can distort ground-motion estimates, particularly peak parameters and spectral amplitudes, which are critical for accurate seismic hazard assessments. As part of our quality control checks, we carefully reviewed the PESMOS dataset and found that the records exhibiting signs of clipping were primarily associated with the vertical component. We sincerely appreciate the reviewer's attention to this issue, as it helped us to revisit and scrutinize the vertical recordings more closely. In response to the reviewer's suggestion, and to support future studies that may require vertical motion data, we have now prepared a separate flatfile specifically for vertical components. While vertical ground motions are indeed important in certain contexts such as near-fault regions or structures particularly sensitive to vertical excitation they were not considered in our current analysis, which is focused on horizontal ground motions. This choice reflects standard practice in ground-motion model development and seismic design, where horizontal components are prioritized due to their dominant role in inducing structural response.

5. The introduction section would benefit from a brief overview of the historical development of strong-motion data compilation efforts in India.

We thank the reviewer for this important suggestion. In response, we have added a brief overview of the historical development of strong-motion data compilation efforts in India in the revised manuscript in lines 89 - 118 "*Such compilation and development of strong-motion data in India has progressed gradually, influenced by both seismic events and institutional initiatives. However, it is worth noting that these efforts were primarily driven by the need to monitor seismic activity, rather than by engineering or design applications. One of the earliest milestones was the installation of modern accelerographs at the Koyna Dam in 1963, which later proved to be critical in capturing the strong ground motions generated by the 1967 Koyna earthquake (Mw ~6.5) a landmark event with first major trigger in India (Tandon and Choudhury, 1968; Chopra and Chakrabarti, 1973).This was followed by the deployment of over 20 RESA instruments and 200 simplified SSR accelerographs by the Department of Earthquake Engineering, University of Roorkee, starting in the 1970s (Rao and Varma, 1981; Rao, 1984). During this period, the Indian Meteorological Department (IMD) and various academic institutions began installing localized instrumentation, particularly in the Himalayan and Indo-Gangetic regions. However, data coverage remained uneven, with dense networks only in select high-risk areas such as Himachal Pradesh, Uttarakhand, and the Shillong region. Recognizing the limitations, the National Strong Motion Instrumentation Project (NSMIP) was launched under the Department of Science and Technology to significantly scale up digital instrumentation. The initiative, led by IIT Roorkee, aimed to install 300 state-of-the-art accelerographs in seismic Zones IV and V, and in densely populated urban centers in Zone III (NSMIP Overview, 2006). Parallel efforts by the National Center for Seismology (NCS), a key agency under the Ministry of Earth Sciences, have contributed significantly to national-level earthquake monitoring and real-time data dissemination. NCS, along with various state agencies, now operates a broad network of seismic stations that complement academic efforts in strong-motion data recording. Additional national programs like the National Programme on Earthquake Engineering Education (NPEEE) and the National Information Centre of Earthquake Engineering (NICEE) have strengthened academic capacity and professional* *training in earthquake engineering (Jain and Agrawal, 2004). Despite these advances, gaps in spatial coverage especially in the central and eastern Himalayan region persist, emphasizing the need for sustained expansion and coordination in strong-motion data infrastructure for accurate seismic hazard assessment. The other important limitations of past efforts include: 1) no standard and uniform quality control criteria to select and reject waveforms, 2) no uniform processing protocols to process the waveforms, 3) the data and the processing details being very rarely disseminated."*

References

Tandon, A. N., & Choudhury, D. C. (1968). The Koyna earthquake of December 11, 1967: Some seismological and geological aspects. Indian Journal of Meteorology and Geophysics, 19(4), 431–439.

Chopra, A. K., & Chakrabarti, P. (1973). The Koyna Earthquake: A review of the seismicity and strong motion records. Bulletin of the Seismological Society of America, 63(6-1), 2063–2080.

Rao, A. S., & Varma, C. V. J. (1981). Strong motion instrumentation in India. Proc. Indian Acad. Sci. (Earth Planet. Sci.), 90(1), 127–134.

Rao, A. S. (1984). Strong motion seismology in India. Proc. Indian Acad. Sci. (Earth Planet. Sci.), 93(1), 127–139.

6. The authors have excluded certain records from the PESMOS dataset. Given the already limited availability of strong-motion data in the Indian subcontinent, it is important to provide a more detailed justification for these exclusions. A transparent explanation such as criteria based on signal quality, metadata incompleteness, or processing issues would enhance the credibility and reproducibility of the dataset.

We thank the reviewer for raising this. We have revised the relevant paragraph in the manuscript to clearly highlight the selection criteria applied to the PESMOS dataset. The rationale for exclusion is detailed under Part II: IGM Processing Protocols (a: Initial Data Quality Checks). In the lines 373-388, we have included the following explanation "*Several records from the PESMOS database were excluded from the present analysis. A subset of recordings was removed due to the absence of clearly identifiable P- and S-wave arrivals, primarily resulting from low signal-to-noise ratios or instrumentation issues. These waveforms, lacking distinguishable phase onsets, were deemed unsuitable for reliable ground-motion parameter extraction and were excluded to maintain the physical robustness and interpretive consistency of the dataset. Additionally, certain events were reported with missing or zero values for key metadata parameters such as magnitude and depth, rendering them invalid for inclusion in ground-motion modeling. Events located at source-to-site distances exceeding 600 km were also excluded, as their recorded ground motions particularly PGA were generally negligible and hold limited* *relevance from a seismic hazard perspective. To ensure regional consistency, the analysis was confined to the Himalayan arc and its immediate foreland basin. Consequently, records from tectonically distinct regions such as Western Pakistan, Afghanistan, and the Andaman Islands were deliberately omitted. These exclusions were made to maintain the geographical and tectonic homogeneity of the dataset, in alignment with the study's objective of developing a region-specific ground-motion model for the Himalaya".*

7. It is recommended to include a discussion on the usable frequency or period range of the ground-motion records.

   We thank the reviewer for the important suggestion. A detailed explanation of the usable frequency range has now been added to the manuscript in lines 439 - 458 "*Also majority of the records were sampled at 200 Hz, giving a Nyquist frequency of 100 Hz. During preprocessing, each record was carefully tapered using a cosine function to reduce spectral leakage, zero-padded to improve filter stability, and then filtered using a 2nd-order high-pass Butterworth filter. Although the filter cutoff frequency was set at 35 Hz in the preprocessing stage to clean up high-frequency content, the Fourier Amplitude Spectra (FAS) were computed over a broad and reliable frequency range of 0.01 Hz to 45 Hz, with a resolution of 0.01 Hz. This frequency band was selected to ensure good signal fidelity, avoid contamination from long-period noise, and stay well within the sensor's bandwidth and sampling limitations. All FAS calculations were performed independently for each horizontal component. The chosen range effectively captures both low-frequency (long-period) and high-frequency (short-period) content relevant for seismic hazard analysis and structural response studies. Frequencies below 0.1 Hz were avoided due to potential contamination from long-period noise and limitations in record duration, while frequencies beyond 45 Hz were excluded due to sensor bandwidth and reduced signal-to-noise ratio. In most cases, the usable low-frequency range falls between 0.1 - 5.0 Hz, while the upper usable frequency varies from 10-45 Hz, depending on the quality and characteristics of the recordings. However, for certain records particularly those from NCS and COSMOS where the usable bandwidth was ambiguous, we conservatively fixed the frequency range to 0.2 - 45 Hz for consistency and to ensure reliable spectral estimation. This defined range ensures the spectral estimates remain stable and interpretable across the dataset".*

8. Please provide more detailed information on how Vs30 values were obtained. Most sources mentioned in the manuscript provide only qualitative soil classifications rather than quantitative Vs30 estimates.

   We thank the reviewer for raising this important point. We fully agree that a clear explanation of how V_{S30} values were obtained is essential for ensuring transparency and consistency in site characterization. Clearly, information of station V_{S30} values is an important issue for such databases as also underlined by the reviewer. However, such information is not easily available and there is no associated documentation available from the network providers. Providing such thorough information is beyond the scope of the current study unless it is made available to us and the wider science fraternity. To clarify this approach more explicitly, we have now added the following lines to the manuscript (lines 756-764) "*Although V_{S30} values are reported for stations, the original source documentation does not provide a detailed explanation of how these values were derived whether through direct measurement, proxy-based inference, or empirical correlations. Since different studies suggest different estimation approaches for different regions in India (e.g., HVSR, MASW, proxy-based), and no uniform methodology is documented across all stations, we have refrained from making assumptions about the V_{S30} derivation. To ensure clarity and avoid misinterpretation, we have included only those V_{S30} values explicitly provided by the network. In addition, we report associated site descriptors such as surface geology and site class, which offer useful context for interpreting site conditions*."

9. The manuscript should explicitly state the limitations of the dataset with respect to Peninsular India, where seismic activity is relatively sparse.

We thank the reviewer for this valuable suggestion. In response, we have rephrased the relevant section of the introduction to incorporate the reviewer's recommendation and to explicitly acknowledge the limitations of the dataset with respect to Peninsular India, where seismic activity is relatively sparse. The following sentences have now been included in the revised manuscript lines 62-77 "*The country spans a geologically diverse landscape, encompassing a range of tectonic environments that give rise to considerable spatial variability in occurrence of earthquakes and thereby in seismic hazard. The Himalayan arc, formed by the active convergence of the Indian and Eurasian plates, is the most seismically active region in the country. As a result, this area has been the primary focus of seismic monitoring and strong-motion instrumentation efforts. In contrast, the Peninsular region, while not entirely free from seismic risk, experiences relatively fewer and more infrequent earthquakes. This lower level of seismic activity has led to limited deployment of seismic instrumentation in these areas, resulting in sparse ground-motion data. The uneven distribution of instrumentation across the country poses challenges for developing a regionally balanced seismic hazard framework. In particular, the lack of sufficient strong-motion records from stable continental regions like Peninsular India limits the ability to fully capture the variability of ground motions across different tectonic domains. Given this disparity, the primary motivation of this study is to develop a uniformly processed ground-motion flat-file, tailored to the Indian context, with a particular focus on the Himalayan region. The decision to focus on this* *area stems from the comparatively higher data availability, which enables more robust analysis and model development".*

10. Contemporary GMMs require comprehensive source parameterization including fault geometry, style of faulting, and rupture directivity effects. The absence of these parameters limits the development of physics-based models necessary for next-generation seismic hazard assessment. Key source characteristics such as fault dip, rake, and strike are absent and should be acknowledged.

    We thank the reviewer for highlighting this important limitation. We fully agree that contemporary Ground Motion Models (GMMs), particularly those developed for next-generation seismic hazard assessments, rely heavily on detailed source characterization such as fault geometry, rupture mechanism, and directivity effects. However, we would like to mention that what parameter (predictor variable) needs to be included in a GMM depends upon the availability of that parameter for example: Ztor (depth to the top of rupture). Hence, it is the database that decides which parameter to include in a GMM not the other way round. Hence, as correctly pointed out, these source parameters (e.g., fault dip, rake, and strike) are not consistently available for most events in the Indian strong-motion datasets. To acknowledge this limitation more explicitly, we have added the following lines to the manuscript lines 771-777 "*The flat-file provides epicentral (R_{Epi}) and hypocentral (R_{Hypo}) distances for all records, computed using the Haversine formula based on the available event and station coordinates. However, due to the lack of detailed fault geometry information such as fault dimensions, dip, strike, and rupture extent it was not possible to compute other commonly used source-to-site distance metrics essential for GMM calibration, such as the Joyner-Boore distance (RJB) and the closest distance to the rupture plane (Rrup). As a result, these distance measures are not included in the current version of the flat-file to avoid inconsistencies and potential inaccuracies".*

11. As the dataset does not include fault mechanism parameters or distance metrics such as R_{JB}, it would be helpful to reference relevant sources or databases that provide finite-fault models.

    We thank the reviewer for this valuable suggestion. In light of the reviewer's recommendation, we have now referenced relevant global and regional sources where finite-fault models are available, which could be beneficial for future efforts involving more comprehensive source characterization. The following lines have been added to the manuscript lines 777-783 "*We thus opted to use only point-source metrics and computing* *distance metrics such as Rrup and RJB was beyond the scope of current study. For events where such parameters are required, researchers may refer to global databases such as the SRCMOD finite-fault model database (Mai and Thingbaijam, 2014) and the USGS Finite-Fault Archive, which provide detailed rupture models for significant global earthquakes. However, the applicability of these databases to Indian events is limited, as very few Indian earthquakes have been characterized with finite-fault models to date.*"

12. Although the authors acknowledge potential uncertainties in metadata, there is no systematic discussion or quantification of these uncertainties, especially for key parameters like hypocentral distance, event magnitude.

    We thank the reviewer for highlighting this point. While we acknowledge that uncertainties in metadata particularly in source parameters such as hypocentral distance and event magnitude can significantly impact the reliability of derived intensity measures, a systematic quantification of these uncertainties is indeed challenging due to limitations in the available documentation and source characterization. In many cases, the distances are based on routine network locations that may be subject to spatial uncertainty, especially in regions with sparse station coverage. Similarly, the magnitude values are typically sourced from different networks, where the methodology (e.g., moment magnitude vs. local magnitude) and associated uncertainties are not consistently reported. To maintain data traceability and transparency, we have retained and reported the original metadata as provided by the contributing agencies. We emphasize that while our dataset represents the most complete and systematically processed collection currently available for the region, users should be mindful of potential variability in source parameter quality. While we fully agree uncertainty in the source parameters is very important for hazard applications, we would also like to mention that such uncertainty estimates in hypocentral depth and magnitude estimates need to be provided by the concerned agencies who prepare earthquake/seismicity catalogues. Clearly, inverting for magnitude and earthquake relocation is beyond the scope of this study. Our future research will be targeted in addressing these challenges.

13. While waveform processing steps are described, the assumptions, filtering choices, and their influence on IM computation are not critically discussed. This is important for reproducibility and transparency.

    We thank the reviewer for raising this important point regarding assumptions, filtering choices, and their influence on IM computation. To clarify, the process implemented in our study is fully automated and based on an adaptive thresholding approach that ensures both objectivity and reproducibility. No fixed cut-off frequencies or manual visual inspections were used at any stage of filtering. Instead, the lower and upper corner frequencies (fc_low and fc_up) were determined individually for each waveform using a consistent algorithm based on the frequency-dependent SNR of the signal. We have made the necessary changes and added within the manuscript, specifically in lines 479-516 *"The noise and signal windows were identified using a P-phase picker algorithm based on the Akaike Information Criterion (AIC), following the approach of Kalkan (2016). This method enabled the automated and consistent detection of seismic phases namely, the noise window, P-wave onset, and S-wave arrival for each waveform. The noise window was defined as the portion of the waveform extending from the beginning of the record up to the P-wave arrival, typically corresponding to the first significant amplitude increase. The signal window was taken as the segment following the P-wave arrival, encompassing the S-wave energy. To prepare the signals for frequency-domain analysis, both windows were subjected to zero-padding to avoid spectral leakage and improve frequency resolution. The resulting amplitude spectra were then smoothed using the Konno and Ohmachi (1998) logarithmic smoothing algorithm with a bandwidth parameter of b = 40, which ensures stability and preserves the key spectral features of the signal and noise. The SNR is calculated by dividing the smoothed root mean square (RMS) amplitude of the signal by the smoothed RMS amplitude of the noise within specific frequency bands. This calculation is performed iteratively over progressively adjusted frequency windows until the SNR exceeds a threshold value of 3. For each frequency window, the mean squared amplitudes of the signal and noise are computed, and their square roots are taken to obtain the respective RMS values. The SNR is then determined as the ratio of the signal RMS to the noise RMS. This procedure is applied independently at both the low and high ends of the frequency spectrum to determine the appropriate cutoff frequencies (fc_low and fc_up). To identify the low-cut frequency, the algorithm begins at a predefined lower bound (e.g., 0.1 Hz) and increases the frequency incrementally (e.g., in 0.1 Hz steps), evaluating the SNR at each step. The lowest frequency at which the SNR surpasses the threshold of 3 is selected as the low-cut frequency (fc_low). Similarly, the high-cut frequency is determined by starting from a higher frequency bound (e.g., 45 Hz) and stepping downward until the SNR again exceeds the threshold, with that frequency recorded as the high-cut frequency (fc_up). This adaptive and automated approach ensures that the filter band is specifically tailored to the signal quality of each individual waveform, allowing for the retention of reliable frequency content while excluding portions dominated by noise. The method avoids reliance on arbitrary fixed bands or visual inspection, thereby improving the objectivity, reproducibility, and robustness of the processing. In particular, by basing the cutoff frequencies on the actual spectral characteristics of the data, this technique enhances the reliability of IM calculations especially at long periods, where low-frequency noise can otherwise introduce significant bias. Inadequate low-frequency filtering may suppress long-period content and underestimate IMs at longer periods, while overestimation of high-frequency noise could bias PGA or short-period SA. By defining the frequency limits* *based on actual signal conditions rather than fixed assumptions, the method ensures that IMs are computed from reliable and physically meaningful portions of the waveform, enhancing both accuracy and scientific credibility"*.

14. Line 250 says 'Vertical-component ground motions were excluded from our present analysis.' where as line 415-417 says "Records having three components (two horizontal and one vertical) are considered. We computed the IMs based on the geometric mean of the two 417 horizontal components". What is the reason behind exclusion of vertical components. Review recommends inclusion of vertical component as well.

    We thank the reviewer for the valuable suggestion regarding the inclusion of vertical-component data. We have updated the manuscript along with the inclusion of the vertical component in a separate flat-file named IGM_flat-file_Vertical. In response, we have revised the manuscript lines 630-640 "*In addition to the two horizontal components, the vertical component of ground motion was also processed following the same waveform analysis and all IMs were computed. This ensures consistency across all three components in terms of filtering, baseline correction, and signal windowing. Although the vertical component was not used in the computation of scalar or rotated horizontal IMs***,** *the resulting vertical IMs have been retained in a separate worksheet within the flat-file for completeness and for potential use in future studies that may require vertical motion data. This approach is consistent with established practices in ground-motion analysis and ensures comparability with other international flat-files and GMM databases. Additional duration measures such as significant duration and Arias duration were computed using the horizontal components to enhance the characterization of ground motion*". to reflect the inclusion of the vertical-component data and to avoid confusion. We appreciate the reviewer's attention to this important detail.

15. Establishing protocols for ongoing database expansion is critical, as isolated database efforts quickly become outdated. A living database structure with clear quality control protocols would significantly enhance long-term utility. The authors should highlight this aspect.

    We thank the reviewer for this insightful and forward-looking suggestion. We fully agree that establishing a living database structure with clear quality control protocols is critical for maintaining the long-term relevance and utility of strong-motion datasets. Indeed, such an approach would allow for ongoing integration of new recordings, updates to metadata, and refinement of derived intensity measures as more information becomes available. However, we would like to note that the present dataset has been compiled using ground-motion records obtained from multiple national agencies. These records were acquired upon formal request and are not directly recorded, owned, or maintained by the authors. As a result, we currently do not have the authority or infrastructure to implement a centralized, automated data ingestion and expansion mechanism. Nonetheless, we recognize the importance of this goal and consider the creation of a community-driven platform in coordination with data-providing agencies a vital next step for the Indian seismic hazard community.

16. It is recommended to maintain a version history and changelog for the dataset. As data or metadata are refined over time (e.g., updated hypocenters, site classifications), users should be able to track changes across versions of the flat-file.

    We thank the reviewer for this important and useful recommendation. As suggested, we would like to highlight that the flat-file, including processed intensity measures and associated metadata, has been uploaded to the Figshare open-access repository. Figshare offers a robust infrastructure for version control, transparent changelog tracking, and clear publication timelines. Each version of the dataset is time-stamped and archived; allowing users to easily view and cite the specific version they are using. We have now also clarified this point in the manuscript lines 1034-1038, emphasizing the benefits of hosting the dataset on Figshare for long-term accessibility, citation, and change tracking. "*Our dataset was posted on July 10, 2025, and any future updates such as refined source parameters, revised metadata (e.g., updated hypocenters, site classifications), or waveform additions will be uploaded as a new version with details of any new updates (related to version) is posted publicity on the portal. This ensures traceability, reproducibility, and continued usability of the dataset for researchers and practitioners".*

17. It is also recommended to Include a mechanism for the research community to suggest corrections, updates, or additions to the dataset.

    We sincerely thank the reviewer for this valuable suggestion. As suggested, we would like to highlight that our dataset has been uploaded to the Figshare repository, which provides built-in mechanisms to facilitate community feedback, collaboration, and version control. Figshare allows registered users to comment directly on uploaded datasets, enabling researchers to suggest corrections, updates, or additions in a transparent and trackable manner. Furthermore, we have now included a statement in the manuscript (lines 1038-1041) highlighting this capability of the repository to encourage open collaboration and continuous improvement of the dataset. "*Figshare supports collaborative project spaces, where creators can invite collaborators who can view, download, and comment on the data offering a practical way for the broader research community to engage with the dataset, propose improvements, and ensure its continued refinement*".

Minor comments

1. The literate citation style the manuscript in inconsistent.

   We thank the reviewer for bringing this to our attention. We have carefully reviewed and revised the manuscript to ensure that all in-text citations and references now follow a consistent literate citation style throughout. This revision improves the clarity and uniformity of the manuscript, and we appreciate the reviewer's helpful suggestion in this regard.

2. There are several grammatical and language-related issues throughout the manuscript. A thorough proofreading or professional language editing is recommended to enhance clarity and readability.

   We sincerely thank the reviewer for this valuable suggestion. In response, we have undertaken a thorough proofreading of the manuscript and revised several sections to improve the grammar, clarity, and overall readability. We believe these changes have significantly enhanced the quality and presentation of the manuscript.

3. Use Repi in place of REpi.

   We thank the reviewer for this suggestion. However, to maintain consistency and clarity across all distance metrics used in the manuscript (such as R_{Epi}, R_{Hypo}, and R_{JB}), we have chosen to retain the current formatting, with capitalized initials representing key components of each term (e.g., E for Epicentral, H for Hypocentral). Moreover, we have clearly defined each of these distance metrics in the manuscript at first use, to avoid any ambiguity for the reader. We believe that this consistent and deliberate use of capitalization across all distance terms helps enhance readability and avoids confusion. We appreciate the reviewer's attention to detail and have ensured that the chosen terminology is applied uniformly throughout the manuscript.

4. Ensure consistent use of symbols and notation across all sections, figures, and equations. Certainly! Here's a polite and clear response to the reviewer:

   We thank the reviewer for this valuable observation. In response, we have carefully reviewed the manuscript to ensure consistent use of all symbols and notations across the text, figures, tables, and equations. This consistency has been maintained in the revised version of the manuscript to improve clarity and coherence throughout. We appreciate the reviewer's suggestion.

5. Conventionally, triangles are used to represent seismic stations. Please revise Figure 2 to reflect this standard practice.

   We thank the reviewer for this helpful suggestion. While we acknowledge that triangles are conventionally used to represent seismic stations, in the revised version of Figure 2, we have used distinct symbols for each seismic network to enhance visual clarity and to allow readers to easily distinguish between different data sources. A legend has been included to clearly indicate the symbol associated with each network.

6. Review equation 5 for typos.
   We thank the reviewer for pointing this out. We have carefully reviewed Equation 5 and rectified the identified error in the revised manuscript.

7. Consider improving the representation of durations in Figure 7 by using vertical lines or markers, which would enhance visual clarity.

   We thank the reviewer for this thoughtful suggestion. In response, we have improved Figure 7 by incorporating vertical lines, markers, and directional arrows to clearly indicate the duration intervals. These enhancements have been made to improve visual clarity and to help readers better interpret the duration measures without ambiguity. We appreciate the reviewer's input, which has helped make the figure more informative and easier to understand.

8. Clearly specify which horizontal component (e.g., geometric mean, maximum, or individual) is used for each intensity measure (IM) reported.

   We thank the reviewer for this important observation. In response, we have clarified the method used to compute intensity measures in the revised manuscript. Specifically, we now state in lines 320 - 325 that *"For our analysis, all intensity measures including peak ground acceleration (PGA), spectral accelerations (SAs), and related parameters were derived using the geometric mean of the two horizontal components, which is a standard approach for representing horizontal ground-motion intensity in a directionally unbiased manner. While the vertical components were also processed using the same methodology, they are provided separately in a dedicated flat-file (IGM_Flat-file_Vertical) and were not included in the primary analyses or visualizations presented in this study."* This revision ensures that the computation approach is clearly and consistently communicated throughout the manuscript.

9. Clarify Time Zone for Event Origin Times.

We thank the reviewer for highlighting this important point. In response, we have now added a column in the flat-file to clearly indicate the time zone information associated with the event origin times for different networks. This addition ensures transparency and facilitates accurate interpretation and comparison of event timing across datasets.

10. Add Flags for Aftershocks vs. Mainshocks if aftershocks are included.

We thank the reviewer for highlighting this point but within the study, we have not applied aftershock declustering when preparing the ground-motion flatfile, following recent trends in seismic hazard modeling where declustering is no longer considered essential. Notably, the 2022 Aotearoa New Zealand National Seismic Hazard Model (NZSHM) developed by Gerstenberger et al. (2023) explicitly avoids declustering, stating that their approach is designed to model the complete seismicity, including aftershocks, foreshocks, and seismic clusters. The rationale is to capture the full spectrum of observed ground shaking, as aftershocks provide valuable data that contribute to hazard estimates and ground motion model calibration. As noted by Gerstenberger et al. (2023), "for the seismicity rate model, aftershock declustering is not required," allowing for a more realistic representation of the seismic environment. Additionally, it is important to recognize that different declustering algorithms such as those by Gardner and Knopoff (1974), Reasenberg (1985), and Zaliapin and Ben-Zion (2013) can produce significantly different classifications of events as mainshocks or aftershocks. This inconsistency introduces uncertainty and variability into the flatfile and can affect ground motion prediction models and hazard estimates. Given this variability and the growing recognition that both mainshocks and aftershocks represent real and relevant shaking, we have retained all recorded events to ensure a comprehensive and representative dataset.

Reviewer A

The earlier observation (comment number 4) *"**The reviewer believes that each of the data are not accessed carefully by the authors. Several of the PESMOS data have suffered from clipping but are included in the manuscript. Such clipped data may demonstrate reliable response spectra and IMs. However, they are not correct and should be excluded. An example is provided below.**"* does not appear to have been adequately addressed in the revised manuscript. Clipping in strong-motion accelerograms occurs when the recorded amplitude exceeds the dynamic range of the sensor, leading to waveform distortion and loss of fidelity in the high-amplitude portion of the signal. These time histories are physically incorrect, and their inclusion can bias parameter estimation, spectral shape, and intensity prediction results. Upon independent verification, I found several ground motion records (***all three components: 2 horizontal and 1 vertical***) used in this study exhibit clear signs of clipping. Such characteristics can significantly affect parameters like PGA, PGV, CAV, and spectral intensity—particularly at shorter periods. These faulty records must be identified and excluded. I strongly recommend that the authors perform a thorough manual review of all accelerograms to eliminate corrupted records from the dataset.

We sincerely thank the reviewer for drawing attention to the issue of clipped PESMOS records. We fully acknowledge that clipped accelerograms can significantly bias the estimation of ground-motion intensity measures, particularly for high-amplitude parameters such as PGA, PGV, CAV, and short-period spectral ordinates. Such distortions, if left uncorrected, can compromise both the accuracy of derived intensity measures and the physical interpretation of ground-motion characteristics. In response to this valuable feedback, we conducted a thorough and systematic re-examination of the entire PESMOS dataset. Each waveform was carefully scrutinized across both horizontal and vertical components to identify any sign of clipping, including subtle cases where the saturation of the recording instrument might not be immediately obvious. During this comprehensive review, we observed that a large fraction of the records obtained after 2014 were affected by clipping. To ensure the scientific integrity of our analysis, all such compromised records were meticulously flagged and excluded from the database.

Following this rigorous screening, we updated the flatfile and revised all subsequent analyses to reflect the corrected dataset. The total number of records, events, and other relevant metadata have been adjusted accordingly throughout the manuscript to maintain consistency and transparency. By restricting our analyses to only reliable, non-clipped records, we have greatly improved the overall quality and consistency of the dataset. This, in turn, strengthens the robustness of the computed intensity measures and ensures that the spectral characteristics reported in our study are both physically credible and free from artifacts introduced by faulty waveforms. These steps not only address the reviewer's concern but also enhance the scientific reliability of the database, providing a more solid foundation for future ground-motion modeling and hazard assessment studies.

We have excluded the records and also added lines to inform about the clipping within the records. Lines 277 - 281 "*Several records were found to be clipped, and all such events were removed after carefully analyzing each component separately. Clipping introduces waveform distortion, which can significantly bias key ground-motion parameters such as PGA, PGV, CAV,* *and spectral intensity particularly at shorter periods. To ensure the reliability of the database, these faulty records were systematically identified and excluded*".

Review B

The revised manuscript is now much improved and incorporates the previous requests for modifications. I only have a few minor additional comments for clarification:

- In the Abstract, the current wording emphasizes an IM database, but the manuscript actually presents a ground-motion flatfile, i.e. a parametric table of IMs and associated metadata. I suggest clarifying this to better reflect the scope and main contribution of the work. Moreover, the reference to "international practice and standards, for example the next generation of attenuation (NGA)" could be misleading, since NGA refers to a U.S.-specific project. To be more general and precise, it would be clearer to state that the flatfile follows practices commonly adopted in the development of Ground Motion Models (GMMs). Indeed, flatfiles represent a widely used international standard to provide uniformly processed IMs and associated metadata, enabling parametric studies, the derivation of GMMs, and other applications in seismic hazard and engineering seismology.

We thank the reviewer for this observation. We have revised the Abstract to clarify that the manuscript presents a ground-motion flatfile (parametric table of IMs and metadata) rather than an IM database. We also replaced the NGA reference with a more general statement that the flatfile follows practices commonly adopted in GMM development. These changes have been made in the Abstract Lines 12 - 17 "*This study presents a ground-motion flat-file for earthquakes recorded in the Indian region, which constitutes uniformly processed intensity measures (IMs) and associated metadata parameters. A major objective of this work is to provide a comprehensive, standardized flat-file that aligns with practices commonly adopted in the development of Ground Motion Models (GMMs). Such flat-files are widely recognized as an international standard, facilitating parametric studies, the derivation of GMMs, and diverse applications in seismic hazard assessment and engineering seismology*".

- In my opinion, **Section 3.2 (Flatfile structure)** would be better placed as an introductory section, replacing the current Part III – Metadata. In fact, this part of the manuscript addresses not only metadata but also IMs, and re-organizing the text accordingly would improve clarity and alignment with the main objective of the work.

We thank the reviewer for this valuable suggestion. We agree that presenting this material as "Flat-file Structure" provides better clarity and aligns more closely with the main objectives of the work. Accordingly, we have carefully revised and reorganized the manuscript: Step III has been retitled from "Metadata" to "Flat-file Structure," and the text within this section has been restructured to reflect this broader scope, which includes both metadata and intensity measures (IMs). The subsequent subsections and references to this part of the manuscript have also been updated to ensure consistency throughout. These revisions were made with the reviewer's feedback in mind and, we believe, substantially improve the readability and logical flow of the paper.

- Please consider titling the sections as *"Step 1, Step 2, …"*. When referring back to them in the text, avoid uppercase letters; for example, line 166, "Step I: Indian Ground Motion (IGM) Database".

We thank the reviewer for this valuable suggestion. We have revised the section titles to and have avoided uppercase Roman numerals. Throughout the text, these are now consistently referred to as "Step I, Step II, …" wherever applicable.

- Correct the subscript notation from *REpi* to *Repi*.

We appreciate the reviewer's suggestion. The subscript notation has been corrected from REpi to Repi, and all corresponding figures and text have been updated accordingly.

- In Table 1, clarify the acronym *MOD* as "Month and Day (of the earthquake occurrence)". For *Site Geology*, please specify according to which classification or site proxy it is defined.

We thank the reviewer for this helpful comment. We have incorporated the requested changes in both the table and the accompanying text. The acronym MOD is now clarified as "Month and Day (of the earthquake occurrence)." For Site Geology, we have added a note explaining that the contributing networks did not specify the methodology or site proxy used in their classifications. Nevertheless, we have incorporated all available geological information from the literature and other relevant sources to provide as complete a description as possible.