doi:10.26443/seismica.v4i2.1620

SEISMICA

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

Shikha Sharma * [1], **Utsav Mannu** [1], **Sanjay Singh Bora** [2]

[1]Department of Earth Sciences, Indian Institute of Technology Gandhinagar, Gujarat, India 382355, [2]GNS Science 1 Fairway Drive, Lower Hutt, 5011 New Zealand

Author contributions: *Conceptualization*: Shikha Sharma, Utsav Mannu, Sanjay Singh Bora. *Formal Analysis*: Shikha Sharma, Sanjay Singh Bora. *Writing - Original draft*: Shikha Sharma. *Writing - review and editing*: Utsav Mannu, Sanjay Singh Bora.

Abstract This study presents a uniformly processed ground-motion flat-file for earthquakes recorded across India, comprising standardized intensity measures (IMs) and associated metadata. The primary objective is to develop a comprehensive, consistent database aligned with international practices for Ground Motion Model (GMM) development. Such standardized flat-files are essential for parametric studies, GMM derivation, and seismic hazard assessment. Thus, we compile a comprehensive dataset of ground-motion IMs, including peak ground motion parameters (acceleration (PGA), velocity (PGV), displacement (PGD)), spectral acceleration (SA), Fourier amplitude spectra (FAS), Effective Amplitude Spectrum (EAS), Arias intensity (AI), cumulative absolute velocity (CAV) significant duration (SD), Acceleration Spectrum Intensity (ASI), Velocity Spectrum Intensity (VSI), and Characteristic Intensity (I_c). These metrics are derived from approximately 778 manually processed strong-motion records from 195 earthquakes, with moment magnitudes (M_w) $\geq$2.0 and epicentral distances (R_{epi}) <600 km, recorded at 254 seismic stations across India between 1986 and 2018. A step-by-step waveform processing protocol was implemented to ensure consistency and accuracy. The reliability of the processed IMs was verified through residual analysis of FAS ordinates against an empirical model. This is the first study of its kind in India to provide a uniformly processed IM database, offering a valuable resource for applications in engineering seismology, seismic hazard analysis, and earthquake engineering.

Production Editor:
Andrea Llenos
Handling Editor:
Carmine Galasso
Copy & Layout Editor:
Kirsty Bayliss

Received:
March 4, 2025
Accepted:
October 10, 2025
Published:
November 14, 2025

1 Introduction

Availability of ground-motion intensity measures (IMs) is crucial for various applications in engineering seismology and earthquake engineering domain. Development of ground-motion models (GMMs), previously known as ground-motion prediction equations, relies on the carefully processed IMs: spectral acceleration (SA), Fourier amplitude spectrum (FAS), and ground-motion duration. The FAS is also important for studies related to estimation of seismological parameters such as stress drop, quality factor, and importantly site amplification. Beyond IMs, the processed waveforms are of crucial importance in earthquake engineering applications, for example time history analysis and selection. Thus, rigorous processing of these ground-motion records, following standardized protocols (Puglia et al., 2018; Luzi et al., 2016), is an indispensable prerequisite for their effective integration into engineering applications. Generally, these processed GM records are compiled, along with associated metadata parameters into a flat-file - a comprehensive parametric dataset that fosters the accurate characterization of ground-motion IMs (Dawood et al., 2016; Bindi et al., 2018). This comprehensive file consolidates event, path, and site metadata along with other seismological parameters (Dawood et al., 2016). Globally, this has been a standard and widely adopted practice for example: the next generation of attenuation (NGA-West1) project mainly focused on California and the western United States (Chiou and Youngs, 2008). A similar coordinated attempt was made for earthquakes recorded in southern Europe and the Mediterranean region in 2011-2014 that culminated in RESORCE database (Akkar et al., 2013). The next stage of the NGA-West2 project also developed a much larger database of ground-motion IMs and associated metadata parameters (Ancheta et al., 2014). Additionally, individual studies also targeted development of such flat-files for ground-motion IMs, including Akkar and Cagnan (2010) for Turkey, Arango et al. (2010) for the Central American subduction zone, Pacor et al. (2011) for the ITACA database in Italy, and more recently, Lanzano et al. (2018) 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. Also, Dawood et al. (2016) developed the first publicly available flat-file of IMs for Japan. Similarly, Bastías and Montalva (2016) developed flat-files of ground-motion IMs for Chile. This database (flat-file) was the basis for development of GMMs that

*Corresponding author: sharma_shikha@iitgn.ac.in

have been used for various applications worldwide.

Within the Indian context over 50% of the land's area is vulnerable to significant seismic activity (Jain, 2016) with high risk to the population centering in the Indo-Gangetic plains (Bilham, 2019). 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.

Despite notable efforts to develop GMMs tailored to Indian tectonic settings, less attention has been paid to the consistent processing of waveform data and systematic documentation of key metadata such as source parameters, source-to-site distances, and site conditions. Moreover, most existing datasets are either not publicly available or lack the transparency required for reproducibility and validation, thereby limiting their utility for broader scientific and engineering applications. In light of these limitations, this study aims to provide a uniformly processed, transparent, and reproducible database of ground-motion intensity measures and associated metadata, with a primary focus on the Himalayan region. The approach ensures that records are selected based on consistent signal quality criteria, and processed using documented protocols. By making this dataset available for further scrutiny and use, we hope to address a critical gap in seismic hazard analysis in the Indian context and enable future efforts toward more accurate and regionally representative ground-motion modeling. The compilation and development of strong-motion data have 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 (M_w ~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 (Mittal et al., 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.

Primarily due to limitations in the available dataset, Sharma et al. (2024) demonstrated that ground-motion characterization in the Himalayan region is associated with considerable aleatory and epistemic uncertainties. These uncertainties are further compounded by the use of varied data processing techniques, which introduce additional epistemic variability in ground-motion predictions. Thus, this paper aims to establish a uniform and processed database platform for the available earthquake datasets in India. It provides a detailed overview of seismic event data collected from various monitoring agencies across the country, outlines the processing protocols applied to Indian ground motion data, and delivers a comprehensive file containing the processed records along with pertinent metadata. The processed number of records and events are rather low in comparison to other globally developed databases and flat-files. However, our aim is to kick-start a convention of compiling the waveforms, developing the flat-files, documenting the processing protocols and making them available for the scientific and engineering community in India for further applications and scientific scrutiny. Finally, the outcome of compiling the waveforms, event, path and site metadata followed by rigorous signal processing protocols are presented in the "Indian Ground Motion (IGM) flat-file". We provide multiple excel sheets within one flat-file (Excel sheet) for various IMs which contain metadata information on events, stations, source-to-site distances, waveforms, and other IMs. The IMs are derived from manually processed recorded waveforms. The IMs provided in the flat-files are peak ground acceleration (PGA, measured in g), peak ground velocity (PGV, measured in m/s), and peak ground displacement (PGD, measured in meters), 5% critically damped SA (measured in g), Arias Intensity (AI), Cumulative Absolute Velocity (CAV), FAS, Effective amplitude spectrum (EAS), measures of significant ground-motion duration, Acceleration Spectrum Intensity (ASI), Velocity Spectrum Intensity (VSI), and Characteristic Intensity (I_c). Additionally, the file incorporates orientation-independent combinations of horizontal components, including RotD0, RotD50 and RotD100 as described by Boore (2010).

Although more recent data may exist, the agencies provided records only up to 2018 at the time we initiated data collection, and hence only these have been included in the database. Hence, we compiled a comprehensive dataset of ~778 strong motion records from 195 earthquakes that were analyzed, incorporating data from 254 seismic stations for moment magnitude M_w ≥ 2.0 and epicentral distance R_{epi} < 600 km recorded within 1986 - 2018 years. This includes: 114 records from 11 earthquakes at 59 stations under the Consortium of Organizations for Strong-Motion Observation Systems (COSMOS) recorded within the years 1986 - 2001 with magnitudes ranging from M_w 4.8 to 7.2 and R_{epi} < 600 km; USGS-NEIC (we hereafter refer to this as aftershocks of the GORKHA earthquake or Gorkha) with 104 records from 18 earthquakes at 8 stations for M_w 5.0 to 7.3 captured within 600 km of the epicentral area; National Center of Seismology (NCS) contributed 62 records from 23 earthquakes recorded at 27 stations between 2016 - 2018 with magnitudes ranging from M_w 4.0 to 6.4 and R_{epi} < 600 km; and PESMOS provided 498 records from 143 earthquake events captured across 160 stations recorded between 2005 - 2014 corresponding to magnitudes M_w 2.3 to 7.8 and R_{epi} < 600 km. For site information, particularly V_{S30}, values provided in the flat-file are directly extracted from the available metadata accompanying the original records, as reported by the respective networks. In most cases, the networks did not specify the methodology used to derive these values whether through direct measurement, proxy-based estimation, or generalized geological assumptions and hence, this information has not been detailed further in the flat-file. Additionally, surface geology classifications were compiled from various literature sources to supplement the site metadata, with the aim of including as much relevant site-related information as possible. In this study, the IGM flat-file is developed as a uniformly processed database of ground-motion IMs and associated metadata. Its structure is organized into six main sections: (i) event-related metadata, (ii) station-related metadata, (iii) source-to-site distance metrics, (iv) waveform-related metadata, (v) IMs, and (vi) additional IMs. A detailed description of each section is provided later, while the manuscript itself is organized into four main steps:

- **Step I Indian Ground Motion (IGM) Database** provides a background on the seismic network utilized in the study and details the strong motion databases from each network.
- **Step II: IGM Processing Protocol** offers an elaborate description of the manual processing scheme applied to the IGM database on each network.
- **Step III: Flat-file structure** outlines the procedure used to compile the earthquake data, distance measures, site characteristics, and other intensity measures for each record. It describes how the flat-file was prepared and explains the derivation of each section within it.
- **Step IV: Additional Ground-Motion Intensity Measures (IMs)** computed in addition to calculating the primary IMs PGA, FAS, and response spectra, we also derived several other IMs of seismological significance, such as AI, duration, and CAV, among others. Further, we conclude with a sensitivity analysis performed on the processed database using the FAS of the IMs.

2 Step I: Indian Ground Motion (IGM) Database

This section highlights the data acquired from different networks either through direct downloads or formal requests. It provides an overview of how each dataset contributed to the analysis, with a detailed breakdown of the earthquake events and corresponding records. Each event is accompanied by detailed information, including the time of occurrence, various magnitude scales provided by different seismic networks, their homogenization to moment magnitude, sampling frequency, as well as the geographic coordinates (latitude and longitude) of both the earthquakes and the recording stations. Additionally, we provide an overview of the seismic networks utilized in this study, alongside a comprehensive summary of the strong motion database compiled from each network. Figure 1 shows the spatial distribution of earthquake epicenters and seismic stations recorded by various Indian seismic networks that were used in our analysis. The dataset covers a geographical extent of 7.9° N-36.5° N latitude and 62° E-95.29° E longitude, encompassing the broader Indian subcontinent. Although the boundaries were defined to include all available data from contributing networks within India and neighbouring regions for regional consistency, the majority of the records are concentrated along the seismically active Himalayan belt, which marks the plate boundary zone.

In this paper, we refer to the time series from individual instrument components as "waveforms", and the combined output of the horizontal components as the

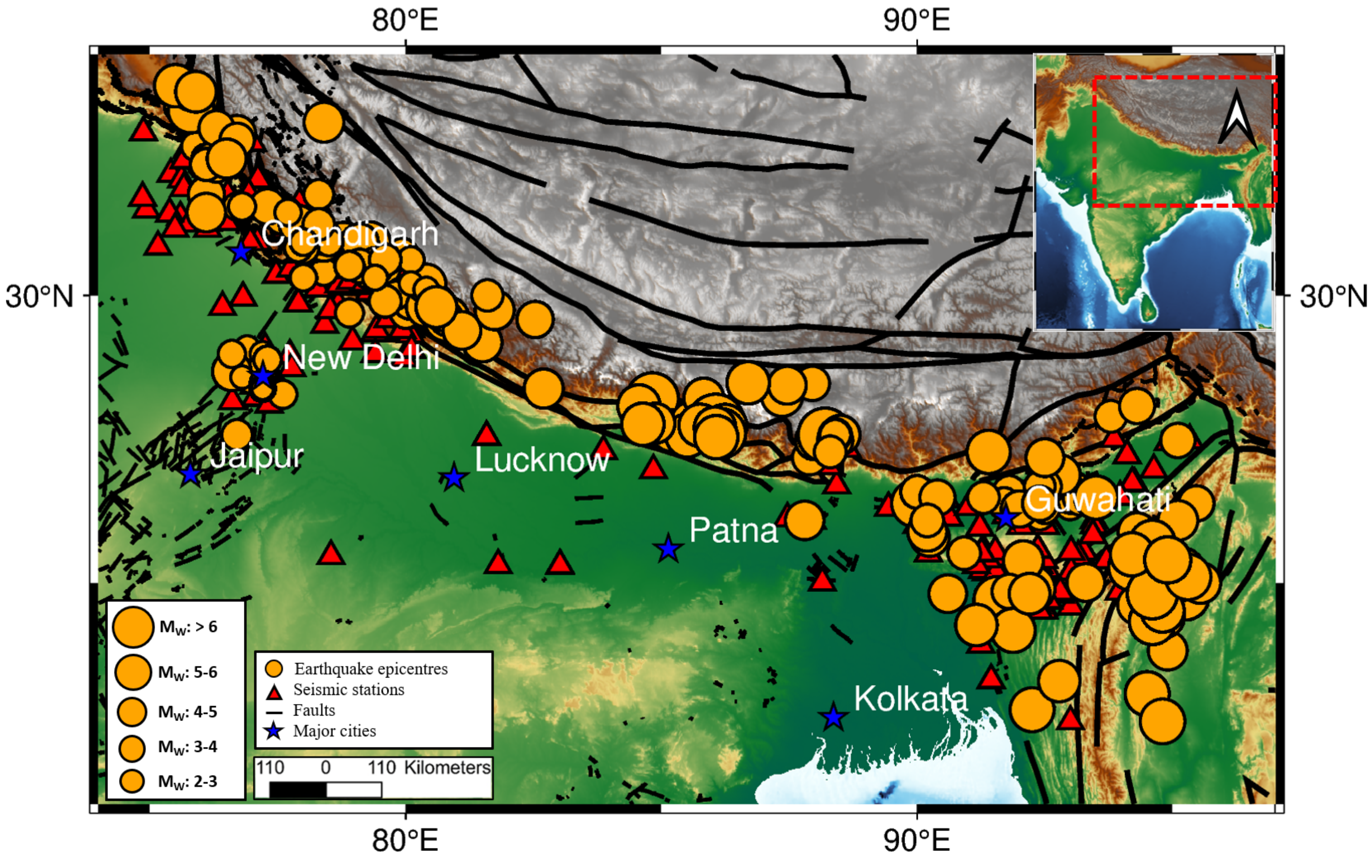

Figure 1 Map showing the epicentral locations of earthquakes (orange circles) with varying diameters of circles representing small to large magnitude earthquakes and seismic stations (red triangles) recorded by various seismic networks. These earthquakes were recorded between 1986 and 2018 by different networks used in our analysis.

record for a specific instrument (e.g., Ancheta et al., 2014). While preparing the database efforts were made to ensure that the processing protocols for all records (from various networks) were consistent. Considering event magnitudes, the contributing seismic networks reported earthquakes using different magnitude scales such as local magnitude (M_L), body-wave magnitude (m_b), and surface-wave magnitude (M_s) as documented in the IGM flat-file. To create a consistent database, the original magnitude types reported by each network were retained in the flat-file, while empirical conversion relationships were applied wherever necessary to obtain moment magnitude (M_w) values for analysis. The compiled dataset includes events with M_w ranging from 2.0 to 7.8. The lower threshold ($M_w \geq 2.0$) was selected from a seismic hazard perspective, as incorporating low-magnitude events improves understanding of attenuation characteristics in regions where larger earthquakes are infrequent. The upper limit corresponds to the largest earthquake recorded within the selected region and time frame. This approach ensures both the completeness and the regional representativeness of the magnitude data used in our analysis. To ensure consistency across the dataset, we converted the magnitudes to a unified moment magnitude (M_w) scale using the conversion relations proposed by Scordilis (2006). We used equations (1) and (2) to convert M_s and m_b to M_w. For the purposes of our current analysis, we treated M_L as equivalent to M_w.

$$M_w = 0.67 \times M_s + 2.07 \quad (1)$$

$$M_w = 0.85 \times m_b + 1.03 \quad (2)$$

a) **Consortium of Organizations for Strong-Motion Observation Systems (COSMOS)** - The strong motion Virtual Data Center (VDC) provides a publicly accessible, web-based search engine for retrieving global earthquake strong-motion data. The desired data was directly downloaded from VDC and used for further processing, requiring minimal preprocessing. For our analysis, we selected 114 records from 11 earthquakes across 59 stations with a sampling frequency of 50 Hz. The selected event data were recorded between the years 1986 - 2001, with M_w ranging from 4.8 to 7.2 and R_{epi} < 600 km. For the majority of the dataset, the original magnitude scale provided was M_s, which was subsequently converted to M_w. Few records had an unknown magnitude type, in which case the original values were retained. The surface geology information provided is categorized sites as either rock or soil. For rock sites, a V_{S30} value of 760 m/s was assumed, while for soil sites, a V_{S30} value of 350 m/s was used. Signal-to-noise ratios (SNR) helped in computing the low and high cut values and were set at around 0.2 and 45 Hz respectively for all the records. In cases where no surface geology information was available, the site was classified as rock.

The VDC was accessed at http://db.cosmos-eq.org (last accessed July 2023).

b) **Aftershocks of the Gorkha earthquake or Gorkha** - The strong-motion stations recorded data from the 2015 Gorkha, Nepal earthquake sequence. A total of 18 mid-to-large earthquake aftershocks comprising 104 records captured at 8 recording stations with M_w 5.0 to 7.3 and within 600 km of the epicentral area were selected. The data files were in ASCII text format with four columns representing time, geographic north-south (NS), east-west (EW), and up-down (UD) components. Each record spans 180 seconds before and after the P and S-wave arrivals to capture the full waveform. The file headers include information on earthquake, site location and recording start time. The datasets required minimal processing, primarily converting the raw WIN system format to acceleration waveform data and applying baseline corrections. The original magnitude types were M_w and m_b (later converted to M_w). All records were sampled at 100 Hz. Site class information was unavailable, so based on the literature (Takai et al., 2016), the sites were classified as either rock or soft soil (lake sediments), with V_{S30} values of 760 m/s and 180 m/s, respectively. The site geology information was taken from the literature (Shigefuji et al., 2022). The data had SNR greater than 3 and similarly this helped in computing the low and high cut values for each record and are documented in the flat-file. Source parameters were obtained from the USGS National Earthquake Information Center (NEIC) https://www.usgs.gov/programs/earthquake-hazards/earthquakes. The data files are archived and accessible via the figshare repository (https://doi.org/10.6084/m9.figshare.19809052, last accessed April 2023).

c) **National Center for Seismology (NCS)** - The governmental agency in India responsible for seismic monitoring, operates the National Seismological Network, consisting of 155 stations nationwide. The data was made available upon formal request after selecting relevant earthquakes from their portal. Each raw datum then underwent rigorous processing using standard protocols (discussed later in Step II) which includes converting the counts, and the signals were normalized to produce acceleration traces. The majority of events were initially measured on the m_b and M_L magnitude scales, which were later converted to M_w. Each earthquake was initially assigned an identification code (e.g., 2016-04-14-0541-39M.OSN_003), which was converted to a Year, Month-Day (MOD), and Hour-minute (HRMN) format. Most events were recorded at a sampling frequency of 100 Hz, with some at 200 Hz. For our analysis, we selected 62 records from 23 earthquakes recorded at 27 stations, recorded for the period 2016 to 2018. These events had M_w ranging from 4.0 to 6.4, and R_{epi} < 600 km. The surface geology data was derived from stratigraphic information found in the literature (Bhandari et al., 1973; Sali, 1970; Kapoor and Singh, 1987; Bansal et al., 2021). The low and high cut values were calculated based on the SNR for each record. These individual values were also determined separately and documented in the flat-files. The data was made available through the portal https://seismo.gov.in/eventwave_datasupply (last accessed August 2022).

d) **PESMOS** - The data was collected and maintained by Indian Institute of Technology Roorkee, utilizing approximately 300 short-period seismometers over the period from 2005 to 2014 (Kumar et al., 2012). Although the data had undergone minimal processing, we carried out a baseline correction or tapering process for a few records to remove residual trends in the signals. 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. Each selected record was processed using a Butterworth filter with a cut-off frequency of 45 Hz. The records had sampling frequencies of 200 Hz, with a few at 100 Hz. The dataset comprises 498 records from 143 earthquake events, captured across 160 stations, with station site classifications as specified by PESMOS. We selected records corresponding to M_w 2.3 to 7.8 and R_{epi} < 600 km. For some records, information about the station name (only station code was available) was unavailable and was therefore left blank in the flat-file. While site geology information was not directly supplied, it was obtained following either through various literature utilizing PESMOS data or the accelerograph stations information provided at https://github.com/anismhd/Pesmos. The V_{S30} values were provided based on the PESMOS site classification, which differs from the NEHRP classification. We computed the low and high cut values based on the SNR for each record. The data was accessed through https://pesmos.org/download-ground-motion-records-2005-2017 (last accessed September 2022).

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. This approach is consistent with previous efforts like the NGA-West2 flat-file, where extended-distance records were also included to improve dataset completeness. Figure 2 below displays the compiled records from all the networks considered in our present study, with M_w ranging from 2.0 to 7.8 and R_{epi} of less than 600 km. The M_w distribution with respect to R_{epi} in Figure 2(a) reveals that a substantial number of records are available for R_{epi} < 400 km, while fewer records exist beyond 300 km. Additionally, only a limited number of records are available for events with M_w < 3 and M_w > 7. Figure 2(b) shows the magnitude and depth distribution, depicting the majority of events occurring within the depth of less than 20 km. Notably, for PESMOS and NCS the depth of records is somewhat fixed to ~10 km (data as provided by different networks). Figure 2(c) presents a histogram illustrating the total number of records and their corresponding earthquakes. 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 uniformf 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.

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.

Data from Gorkha, COSMOS and PESMOS required minimal adjustments during processing to generate the desired intensity measures (IMs). In contrast, data from NCS underwent comprehensive processing using standard protocols (outlined in the subsection) to maintain consistency with methods used by other seismic networks, ensuring homogeneity across the dataset. For our analysis, we used the geometric mean of the two horizontal components for PGA, PGV, PGD, and SA, while the horizontal components were used separately when computing the other IMs (e.g., FAS, rotation-independent measures), 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. Figure 4 presents the distribution of IMs as a function of magnitude and hypocentral distance for the compiled dataset. Figure 4(a) shows variation of PGA with moment magnitude (M_w), while Figure 4(b) illustrates the expected attenuation of PGA with increasing epicentral distance (R_{epi}). Figure 4(c) and (d) display the variation of SA at 0.5s and 1.0s, 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.

3 Step II: IGM Processing Protocol

This section outlines the detailed manual processing protocol applied to the IGM database. The available records were analyzed for 2 minutes before and 5 minutes after the event origin time to capture the entire signal ensuring that the full usable waveform was correctly identified. 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, 2010). 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. For the records obtained from COSMOS, the signal duration was very short; therefore, the entire signal duration was analyzed. We also filtered out high- and low-frequency noise from the records and corrected the waveforms to eliminate any trends. Additionally, the instrument response was also deconvolved. The final acceleration and velocity values were set to zero which is crucial for accurate IM calculation. 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 included with little processing using the same methodology in the flat-file as a separate worksheet for completeness and future use.

The following steps were undertaken to generate the processed waveforms from each record in our present analysis:

a) **Initial data quality check** - Each record underwent rigorous quality checks before being incorporated into the flat-file to ensure the reliability of the dataset. This process involved a thorough evaluation of waveform amplitudes to identify and exclude records with anomalous ground motion intensities, often caused by erroneous response files. Data from COSMOS required minimal processing, with only baseline corrections applied to a few

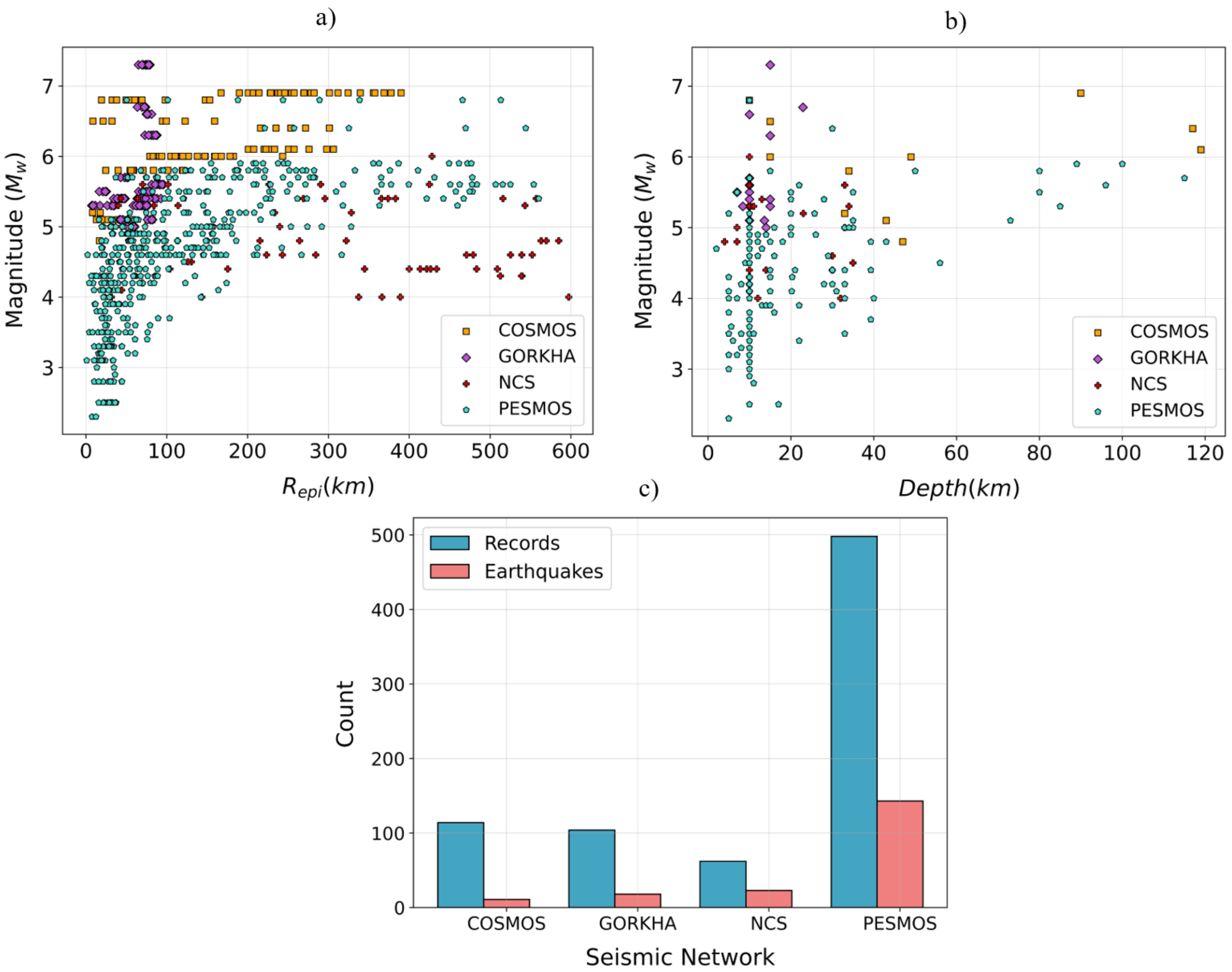

Figure 2 (a) Distribution of moment magnitude (M_w) versus epicentral distance (R_{epi}) for the records used in our analysis, (b) Moment magnitude (M_w) versus depth distribution of the events, and (c) Histogram displaying the total number of records and associated earthquakes sourced from various seismic networks.

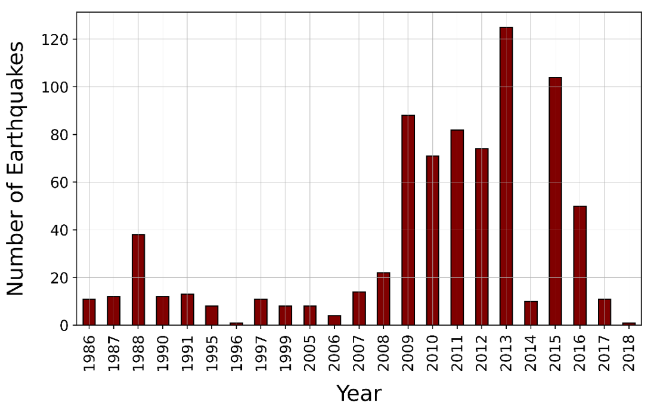

Figure 3 Annual Distribution of Earthquake Records (1986 - 2018)

records. In contrast, records from NCS underwent extensive processing 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. The records from Gorkha also required minimal processing as they had already been pre-processed before inclusion. 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

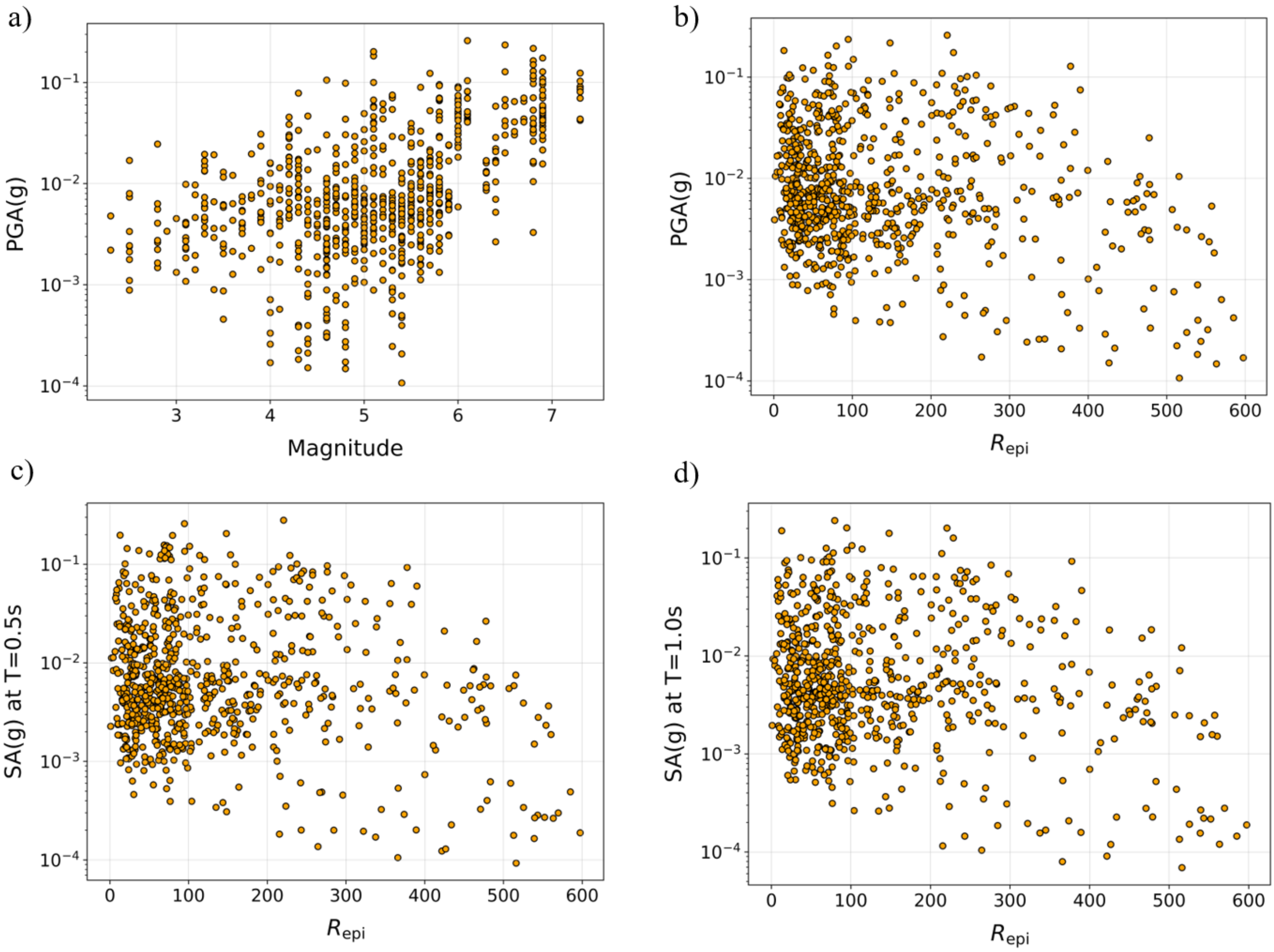

Figure 4 Magnitude and distance scaling of Peak Ground Acceleration (PGA) with (a) moment magnitude and (b) epicentral distance, and Spectral Acceleration (SA) with distance at (c) T = 0.5s and (d) T = 1.0s.

their recorded ground motions particularly PGA were generally negligible and hold limited relevance from a seismic hazard perspective. A few clipped records were also excluded as already discussed. 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 objective of developing a region-specific ground-motion model for the Himalaya. Some records underwent additional steps, including baseline corrections and detrending, to remove any trends present in the acceleration traces. This step was crucial to ensure accurate and realistic measurements.

Figures 5 (a) and (b) present examples of waveforms from the PESMOS network that were deemed acceptable, confirming the clear identification of distinct P or S-waves. In contrast, Figures 5 (c) and (d) display records from the NCS network that were rejected due to the absence of distinguishable features. This step was crucial in determining the usable frequency range and ensuring the preservation of the earthquake signal's inherent characteristics. 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.

b) **Removal of instrument response** - The instrument response was meticulously removed using pole-zero files supplied by the respective agency. The ObsPy tool (Beyreuther et al., 2010) was employed to convert all waveforms into acceleration time series, ensuring consistency across the dataset. Subsequently, the acceleration data was normalized to standardize the amplitude across different events and stations, facilitating uniform and accurate analysis for ground motion studies. This

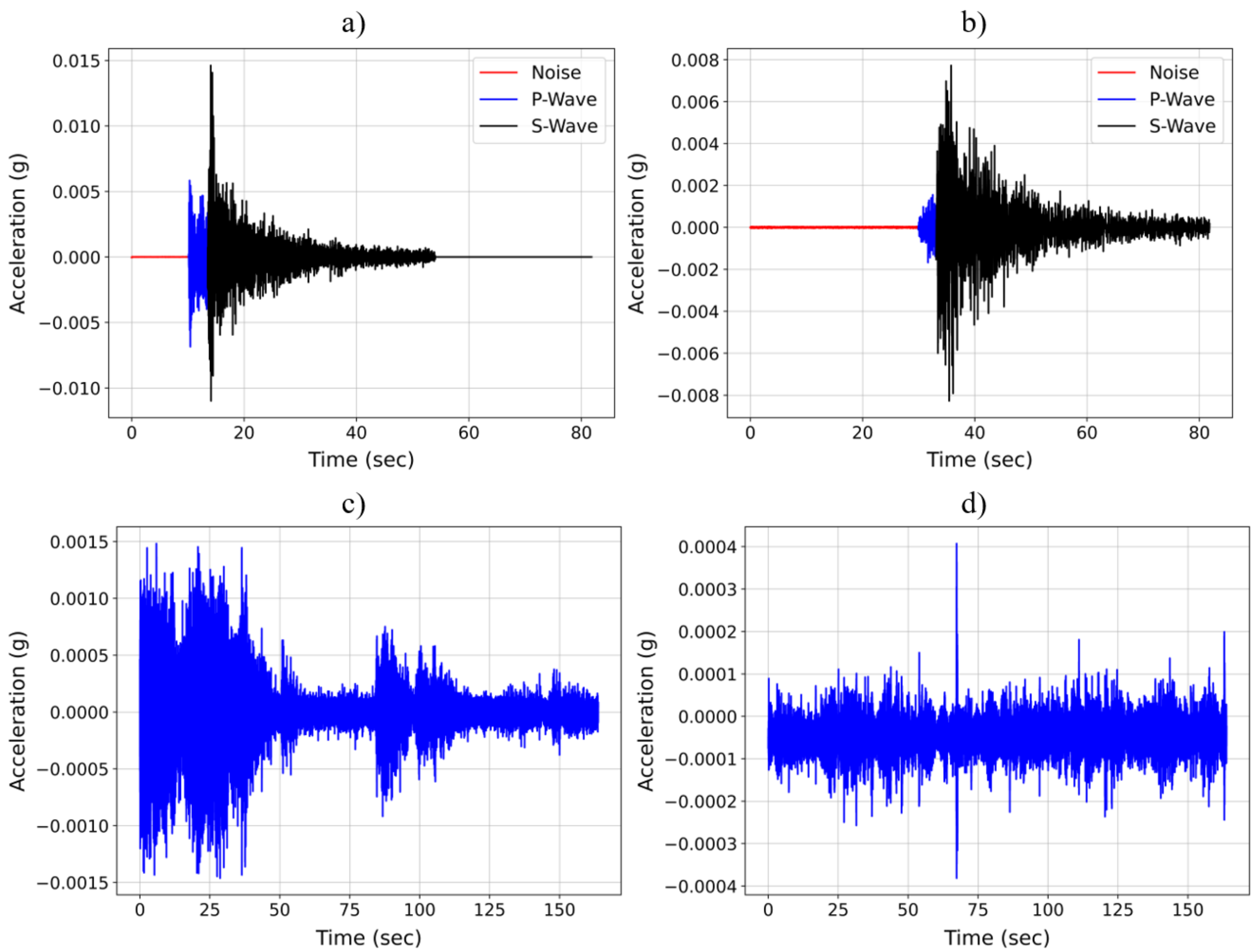

Figure 5 (a) and (b) Examples of acceptable waveforms from PESMOS records, showcasing clear noise levels and distinct P- and S-wave arrivals. (c) and (d) Acceleration plots highlighting reasons for rejecting certain records from the NCS database.

step was crucial for maintaining consistency in the dataset and was especially required for NCS records, since it had already been carried out for records from other networks. Many NCS records required extensive processing due to the lack of clear P and S-wave arrivals. In contrast, the records from other networks, including COSMOS, Gorkha, and PESMOS, required minimal processing, as previously discussed. Such datasets were already preprocessed or contained sufficient metadata, rendering additional steps like instrument response removal unnecessary. For instance, COSMOS and Gorkha records were either already corrected or processed to a level where no further adjustments were required. Similarly, the PESMOS records underwent basic baseline correction and detrending, which addressed the key issues without requiring the removal of the instrument response. This variation in processing reflects the differing levels of quality and completeness inherent in the datasets. 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., 2025). 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.

c) **Windowed signals** - Each record was meticulously windowed to capture all potential seismic phases, while incorporating sufficient pre- and post-event noise to ensure signal quality. Depending on the variation in record durations, we applied appropriate time windows, such as 120 or 170 seconds, tailored to each record depending on the sampling frequency. Corresponding bandpass filtering ranges were also applied to enhance the SNR. In cases where pre-event noise was unavailable, the windowing was adjusted so that the earliest available time marked the start of the window, preserving the integrity of the seismic data and critical signal characteristics across the dataset. Also, the 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 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.

d) **Baseline correction** - A zero-order baseline correction was applied by subtracting the mean value from the entire waveform to eliminate persistent low-frequency, nonphysical trends in the time series after filtering. The acceleration time series of each record was subsequently converted to velocity, and the effectiveness of the record processing was validated by ensuring that the ground velocity returned to zero at the end of the earthquake-induced shaking. The correction ensures that the long-period noise or drift is removed from the data, enabling more accurate representation of the seismic signal so that the data retains its physical significance, allowing for a more reliable analysis of ground motion intensity and waveform features. In our dataset, one or more baselines (either straight lines or low-order polynomials), were subtracted from the acceleration trace, as outlined by Boore (2005).

e) **Tapering and padding** - A cosine taper window was applied to the beginning and end of the signal to ensure a smooth transition to zero, with the taper length set at 5% of the total record duration. Additionally, zero padding was also added before and after the acceleration time series, in accordance with the guidelines provided by Converse and Brady (1992). This step is essential to avoid distortion in the filtered motion, as noted by Boore (2005). In some cases, additional zero padding was required to adjust the record length to a power of 2, a prerequisite for precise Fast Fourier Transform (FFT) calculations.

f) **Signal-to-noise ratio (SNR) analysis** - 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.

Figure 6 illustrates the analysis of an earthquake signal recorded during the Pithoragarh earthquake (M_w 4.3) recorded at a distance of 26 km from PESMOS network. Subplots 6(a) and (b) display the acceleration time series for the two horizontal components, H1 and H2, respectively. In both plots, the noise window (shown in red) spans the initial 14.8 seconds before the P-wave arrival. Subplot 6(c) shows the FAS for the corresponding raw signal and noise components across the entire frequency range. The raw FAS values are plotted across the full frequency spectrum, while the FAS for the noise is restricted to the relevant frequency range under consideration. The frequency analysis was performed systematically across all records to ensure data quality prior to further processing. This assesses the quality of the seismic records and removes background noise to enhance the accuracy of the record.

g) **Waveform filtering** - In our study, we applied causal Butterworth bandpass filters to all 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, 2010). 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.

All processed acceleration time histories, following above processing steps, were saved for further analysis. It is important to note that the manual processing protocol discussed in this manuscript is tailored specifically for strong motion datasets used by the seismic network discussed above (Step II). Additional checks and processing steps may be required for broadband or analog instruments. For more detailed guidance on processing analog seismic records, readers are encouraged to refer to Douglas (2003) and Boore and Bommer (2005).

4 Step III: Flat-file structure

The parametric file includes detailed earthquake data, such as magnitude, epicenter and hypocenter distances, source characteristics, site characteristics including geology, and shear wave velocity (wherever provided by the network), from manually processed waveforms. We present an overview of the IGM flat-file structure, accompanied by a flowchart that illustrates the simplified organization and later the compilation process of each metadata component. The fields in the flat-file are systematically organized into five main metadata categories, following the widely adopted structure proposed by Lanzano et al. (2018). This categorization ensures that all essential seismological parameters are coherently captured and integrated within the IGM database. To aid in the understanding of this structure, Figure 7 presents a detailed flowchart outlining the categorization, which comprises: (i) Event-related metadata, (ii) Station-related metadata, (iii) Source-to-site distances, (iv) Waveform-related metadata, (v) IMs and (vi) Additional IMs.

Event-related metadata: This includes essential parameters that define each seismic event. Specifically, it comprises: (i) unique identification codes for each event and corresponding record, along with the origin time of the event; (ii) event location parameters, including geographic coordinates (latitude and longitude) and focal depth; and (iii) various magnitude estimates, such as moment magnitude (M_w), local magnitude (M_L), surface-wave magnitude (M_S), and body-wave magnitude (m_b), as detailed in the IGM database documentation (Step I). While a subset of the events originally had M_w reported by the contributing agencies, many others were recorded using different magnitude scales (M_L, M_S, or m_b). To maintain consistency across the dataset in ground-motion modeling, all magnitude values were homogenized to Moment Magnitude (M_w). This conversion was performed using established empirical relationships, as described in Equations 1 and 2, ensuring a standardized magnitude representation throughout the database. This uniformity is crucial for accurate source characterization and consistent interpretation in subsequent ground-motion studies.

Station-related metadata: This section includes detailed parameters associated with each recording station. Key fields provided in the flat-file include the station identification code, station coordinates (latitude,

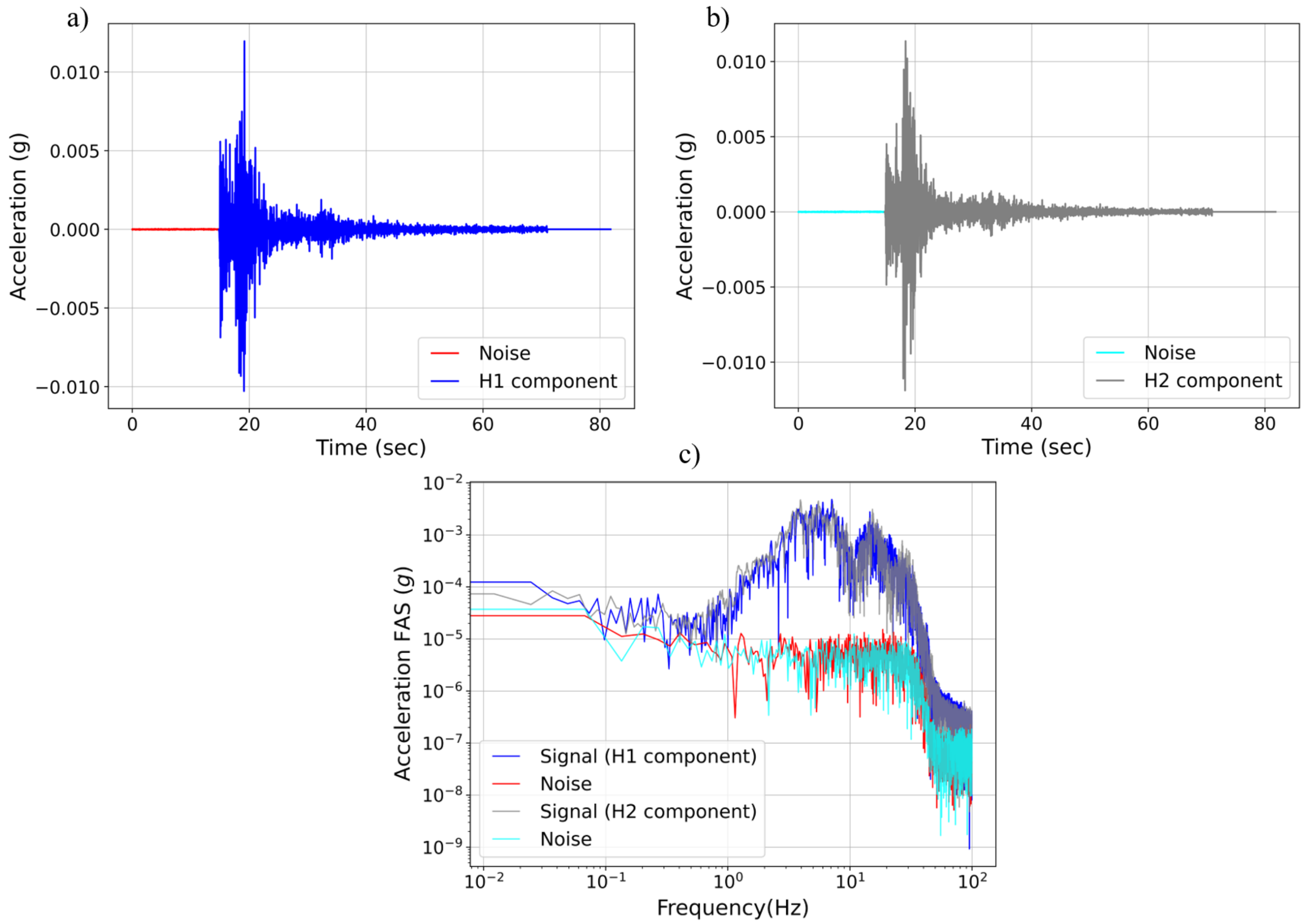

Figure 6 Acceleration time series for the horizontal (a) H1 and (b) H2 components along with the noise window (~14.8 sec) for Pithoragarh earthquake from the PESMOS record with M_w 4.3 recorded at a distance, R_{epi} 26 km and (c) Fourier Amplitude Spectrum (FAS) plot displays raw FAS values for both the signal and noise across the entire frequency range of the record.

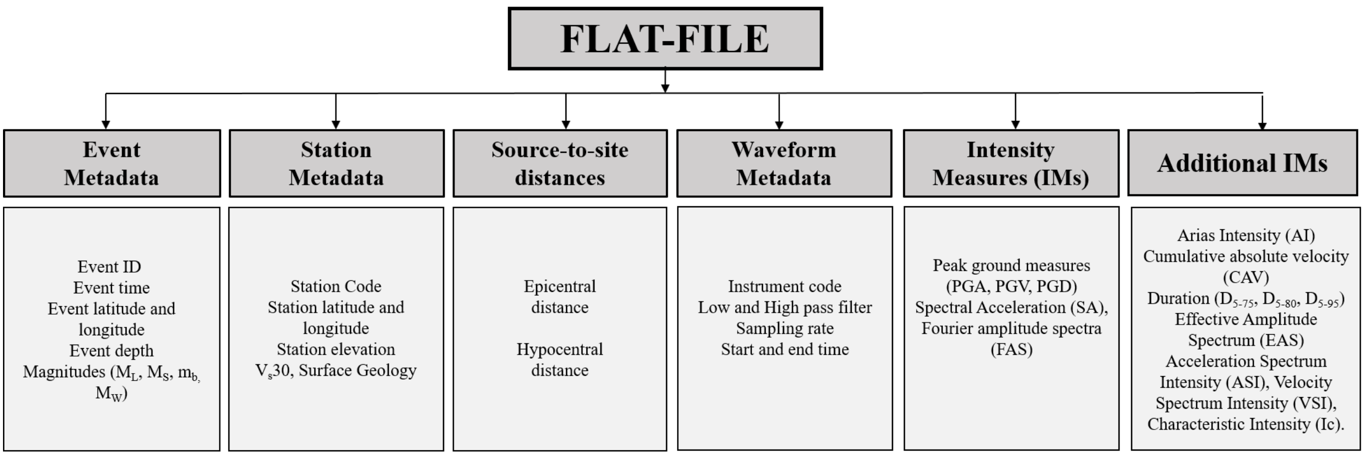

Figure 7 Schematic structure of the IGM flat file contents

longitude, and elevation), and available site condition metadata. Site classification is presented in two parts: (i) the classification assigned by individual networks, particularly for PESMOS data, which follows a scheme distinct from NEHRP standards (with further details documented in the IGM_Flat-file), and (ii) the NEHRP site classification, wherever sufficient V_{S30} data is available. Additionally, surface geology information was compiled from available network downloads or relevant published literature, allowing for a broader understanding of site conditions. 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. For certain stations, due to incomplete metadata availability, only the station code is provided while station name, V_{S30}, or surface geology fields have been left blank where this information could not be reliably obtained. These fields are clearly marked, ensuring transparency and enabling future updates as more data become accessible. 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.

Source-to-site distances: 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 (R_{JB}) and the closest distance to the rupture plane (R_{rup}). As a result, these distance measures are not included in the current version of the flat-file to avoid inconsistencies and potential inaccuracies. We thus opted to use only point-source metrics and computing distance metrics such as R_{rup} and R_{JB} 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.

Waveform metadata: 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 Step 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.

Intensity Measures (IMs): This includes all associated metadata along with calculated peak values: PGA, PGV, and PGD. PGA was calculated as the geometric mean of the two horizontal components. PGV and PGD were calculated separately for each horizontal component using the cumulative trapezoidal integration method, after which their geometric mean was computed to obtain the final values. Additionally, SA values were computed for a range of oscillator periods from 0.01 to 10 seconds for 37 periods using a 5% critical damping ratio, following conventional practice in seismic hazard analysis and GMM development. These IMs provide a robust basis for characterizing ground motion behavior across a wide range of structural response periods and are essential for applications in engineering design, risk assessment, and empirical modeling. FAS was computed for each record to capture the frequency-domain content of ground motion, using the Konno and Ohmachi (1998) smoothing method with a bandwidth parameter b = 40, which balances spectral resolution and noise reduction. FAS was calculated separately for the two horizontal components (H1 and H2) to preserve directional frequency characteristics, providing spectral information that complements time-domain IMs and supports site response analyses and comparison with theoretical models. Further, to account for orientation effects, we applied the Boore (2010) method to determine rotation independent measures of IMs (i.e., RotD0, RotD50 and RotD100) over the period range of 0.1 to 10 seconds. RotD0 (rotational average of the 0^{th} percentile of the peak ground motion) refers to the minimum rotated component of a ground motion time series. It provides a conservative, angle-independent measure of ground shaking, representing the least extreme amplitude recorded when horizontal components are rotated through all possible orientations. RotD50 (rotational average of the 50^{th} percentile of the peak GM) refers to the median rotated component of a ground motion time series and gives a representative, angle-independent measure of the ground shaking. RotD100 (rotational average of the 100^{th} percentile of the peak GM), on the other hand, refers to the maximum rotated component of a ground motion. To compute these, the longitudinal and transverse components of ground motion were first rotated to obtain two orthogonal horizontal components: H1 and H2. These components were combined into a single time series, which was then rotated through angles ranging from 0° to 180° at 1° increments. The rotated horizontal component at any angle θ was computed using the following equation:

$$H_{\mathrm{ROT}}(t;) = H1(t)cos(\theta) + H2(t)sin(\theta) \quad (3)$$

where H_{ROT} represents the horizontal ground motion at time t and angle θ, while H1 and H2 are the orthogonal components of ground motion.

Figure 8 shows the comparison of SA values for two different earthquake scenarios, their individual components, their geometric mean, RotD0, RotD50 and

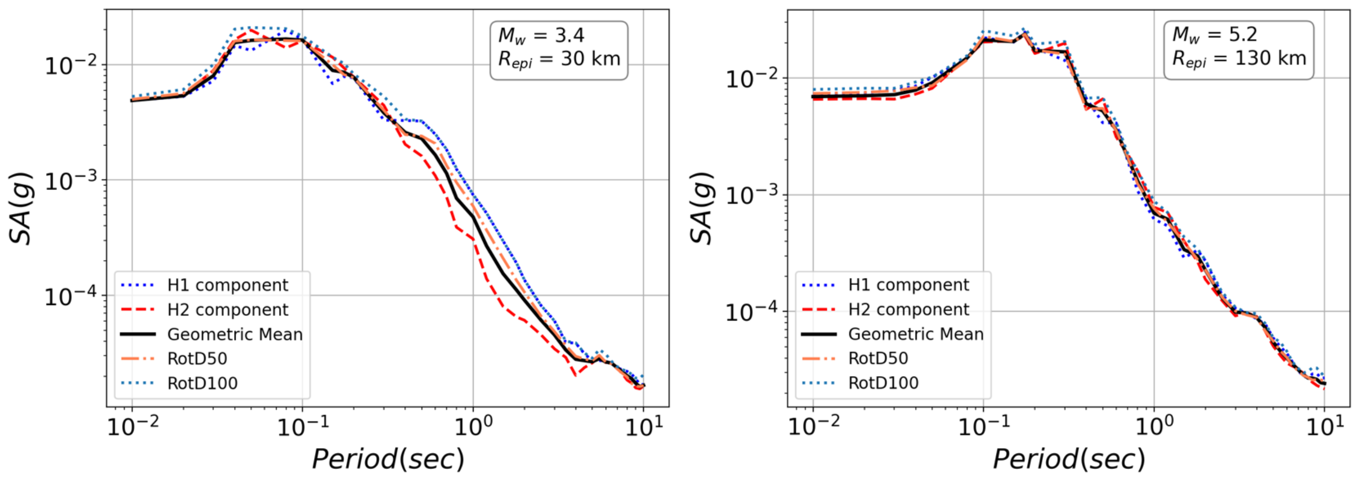

Figure 8 Comparison of Horizontal Ground Motion Intensity Measures (Geometric Mean, RotD50, RotD100) for two different records using PESMOS data.

RotD100 variation. The data is utilized from the PESMOS network with magnitudes M_w 3.4 and 5.2 and epicentral distances R_{epi} = 30 km and 130 km. In the left panel one can observe significant differences (mainly at longer periods) between the spectra for individual components and the combined spectra. In the right panel for a moderate magnitude event at a larger distance such differences are not observed. Clearly, any meaningful inferences cannot be made from these plots regarding differences between various component definitions. However, they are shown to provide visual sanity checks.

Additional Intensity Measures (IMs): Apart from the primary IMs, we also computed several advanced measures to capture the energy, duration, and frequency characteristics of strong ground motions. These include significant duration metrics based on the accumulation of AI (Trifunac and Brady, 1975; Kempton and Stewart, 2006), EAS (Goulet et al., 2018; Manea et al., 2023), ASI (Von Thun, 1988), VSI (Housner, 1959), and I_c (Park and Ang, 1985; Liao et al., 2022; Sriwastav et al., 2024; Dai et al., 2023). These parameters, described in detail in Step IV, are included in the flat-file to provide a comprehensive characterization of ground-motion amplitude, energy content, and frequency-dependent features, supporting a wide range of applications in engineering seismology and seismic hazard assessment.

Consistent with standard GMM practices, most IMs were computed using the geometric mean of the two horizontal components, while a few were derived from the individual horizontal record when required. The vertical component was also processed following the same waveform analysis procedures (filtering, baseline correction, and signal windowing) to ensure consistency across all three components. Although the vertical motions were not used in the computation of scalar or rotated horizontal IMs, the resulting vertical IMs are included in a separate worksheet of the flat-file for completeness and to facilitate future studies that may require vertical ground-motion data. This approach aligns with established protocols in ground-motion analysis and supports direct comparability with other international flat-files and GMM databases. We organized the dataset into separate excel sheets to streamline the dissemination of our findings. This approach helped to manage the data more effectively and made it easier to share. By splitting the data, we avoided creating an excessively large table that might be difficult to handle using spreadsheets. This methodology aligns with best practices, as suggested by Lanzano et al. (2018). These files are provided under the name "IGM_Flat-file.xlsx" and include multiple sheets within a single file, each containing specific datasets. For vertical component "IGM_Flat-file_Vertical.xlsx" is provided. Comprehensive metadata for each considered earthquake was compiled, including details such as event magnitude, location, and source characteristics. This information provides the foundational context for all subsequent analyses. Further details of the flat-file, including the acronyms and header names used to describe key seismological parameters related to the source, path, and site along with their definitions are provided in Table 1. The table highlights the most important metrics used to characterize these properties, offering a concise summary of the key parameters contained in the flat-file.

5 Step IV: Additional Ground-Motion Intensity Measures (IMs)

Empirical GMMs have also been developed for cumulative and spectrum-based measures of earthquake shaking, including AI, CAV (Campbell and Bozorgnia, 2010; Abrahamson et al., 2016; Campbell and Bozorgnia, 2019; Farhadi and Pezeshk, 2020; Stafford et al., 2008), and significant duration (Kempton and Stewart, 2006; Bommer et al., 2009). To evaluate the temporal evolution of shaking, we used Husid plots (Bommer and Martínez-Pereira, 2000) to track the cumulative AI over time, allowing estimation of significant duration as the time interval between specified AI thresholds. In addition to these cumulative and duration-based metrics, we also computed other intensity measures to capture the frequency content and energy of ground motions: EAS, ASI, VSI, and I_c. These complementary IMs provide a more complete characterization of strong ground motion.

a) **Computation of AI, CAV, and significant duration**

- **Arias Intensity (AI)** - A duration-related parameter that measures the total earthquake input energy per unit weight for a series of undamped elastic single-degree-of-freedom systems with frequencies uniformly distributed from zero to infinity (Arias, 1970; Bozorgnia and Campbell, 2004). It is defined as the integral of the squared acceleration time series (Arias, 1970; Benjamin, 1988) and can be expressed as follows:

$$AI = \frac{\pi}{2g} \int_0^T a(t)^2 dt \quad (4)$$

where $a(t)$ is the ground acceleration in m/s^2, g is the gravitational acceleration, T is the total duration of the ground motion recording, and AI is in m/s. AI was computed using the geometric mean of the acceleration traces, processed separately for records with different sampling frequencies. To compute AI, the geometric mean (initially in g for all records) was converted to m/s^2. While the time window for calculating AI depended on the total length of the waveform and sampling frequency, it was limited to 0 - 80 seconds for inclusion in the flat-file. AI quantifies the total seismic energy imparted to a structure during an earthquake over time highlighting the relative timing of the energy release, with a rapid rise indicating a short burst of intense shaking.

- **Cumulative Absolute Velocity (CAV)** - A crucial parameter that quantifies the total area under the absolute ground acceleration curve, $|a(t)|$, over the duration of the shaking. Defined as the integral of the absolute acceleration time series, CAV provides a measure of the cumulative effect of ground shaking, reflecting the total velocity change experienced by a structure during an earthquake. Mathematically, CAV is calculated as:

$$CAV = \int_0^T |a(t)| dt \quad (5)$$

where $|a(t)|$ is the absolute acceleration with respect to time (m/s^2), and T is the total duration of shaking. CAV accounts for the cumulative effects of ground motion duration and was computed, similar to AI, using the geometric mean of the two horizontal components. The computation was limited to a 0-80 second time window for inclusion in the flat-file. It captures the integrated impact of shaking over time, correlates strongly with observed structural damage, and serves as an effective predictor of ground motion intensity in structural assessments (Campbell and Bozorgnia, 2006).

- **Significant Durations** - Ground-motion duration was quantified using significant duration metrics, defined as the time elapsed between specified levels of cumulative AI (Arias, 1970; Bommer and Martínez-Pereira, 2000). These metrics capture the interval over which most of the seismic energy is released and are widely used in engineering seismology because of their relevance to structural response and potential damage. Several predictive models for significant duration have been developed in engineering contexts (e.g. Trifunac and Brady, 1975; Dobry et al., 1978; Kamiyama, 1984; Kempton and Stewart, 2006), providing a well-established framework for assessing the energy-related characteristics of strong ground motion. Husid or $H(t)$ plots, are used to illustrate the cumulative normalized energy, often represented by AI, as a function of time. This energy is expressed as a percentage of the total energy. Generally, the time intervals between specific thresholds of cumulative AI values: $H(t) = 5\%$ and 75% (denoted as $D_{5\text{-}75}$), $H(t) = 20\%$ and 80% ($D_{20\text{-}80}$) and $H(t) = 5\%$ and 95% ($D_{5\text{-}95}$) are considered providing valuable insights into the earthquake's energy distribution and duration characteristics.
  - $D_{5\text{-}75}$: Time interval between the instants when 5% and 75% of the total AI is reached, representing the period over which the central 70% of the seismic energy is released and primarily capturing the energy associated with body waves.
  - $D_{20\text{-}80}$: Time interval between 20% and 80% of cumulative AI, focusing on the window of peak energy release and providing an intermediate measure that includes both body-wave energy and some contributions from surface waves.
  - $D_{5\text{-}95}$: Time interval between 5% and 95% of cumulative AI, encompassing nearly the entire duration of significant shaking and often used in dynamic structural analyses, as it captures low-frequency surface waves and coda energy (Bommer et al., 2009; Meimandi-Parizi et al., 2020).

These duration measures are calculated from the cumulative AI 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). Figure 9 illustrates the significant duration calculated for the three matrices, demonstrating how each matrix captures the total duration of the effective signal during an earthquake. The plot reveals how different matrices account for the time intervals between various cumulative energy thresholds (e.g., $H(t) = 5\%$ and 75%, $H(t) = 20\%$ and 80%, and $H(t) = 5\%$ and 95%) providing a comprehensive view of the earthquake's energy distribution and the duration of its effective signal.

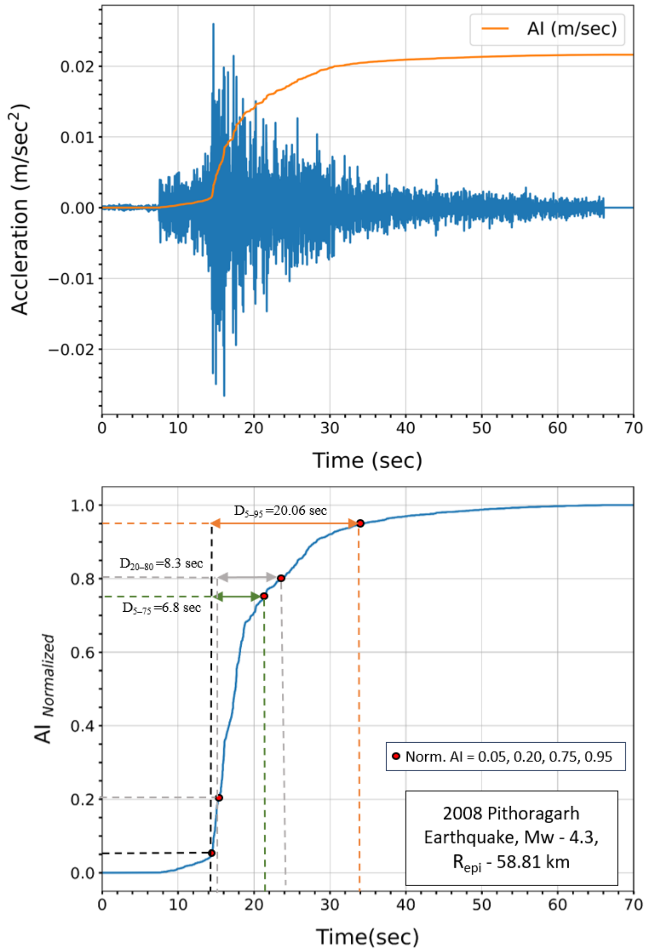

Figure 9 The figure shows a plot of a recorded ground motion of one record (above) and Husid plot (below) with significant durations (D_{5-75}, D_{20-80} and D_{5-95}).

b) **Effective Amplitude Spectrum (EAS) -** In addition to the FAS from the two individual horizontal components, we provide a flat-file for EAS. It combines the smoothed FAS of the two orthogonal horizontal components and is considered equivalent to rotation independent measure of the combined FAS from the two components. EAS is calculated for an orthogonal pair of FAS using the following equation (Goulet et al., 2018; Manea et al., 2023):

$$EAS(f) = \sqrt{\frac{1}{2}\left(FAS_{H1}^2 + FAS_{H2}^2\right)} \tag{6}$$

where, FAS_{H1} and FAS_{H2} are the FAS of the two recorded orthogonal horizontal components of the ground motion and f is the frequency in hertz. This method helps to capture the overall amplitude characteristics of ground motion, making it particularly useful for engineering applications where orientation-independent measures are desired. Figure 10 illustrates the FAS for two horizontal components, FAS_{H1} and FAS_{H2}, along with their respective smoothed curves using Konno-Ohmachi smoothing (b = 40) for 2005 Chamoli earthquake (PESMOS database) at 100 frequency ordinates in the range 0.1-30.0 Hz. The top left plot shows the unsmoothed and smoothed FAS_{H1}, while the top right displays the same for FAS_{H2}. The bottom plot shows the EAS derived from both smoothed FAS of the two horizontal components, and further compared with the smoothed versions of FAS_{H1} and FAS_{H2}. The smoothed curves provide a clearer representation of the overall spectral trend, reducing the ruggedness seen in the raw amplitude spectra.

c) **Acceleration Spectrum Intensity (ASI)** - The integral of the PSA over the period range 0.1- 0.5 s (Von Thun, 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:

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

Where $PSA\,(\zeta = 0.05, T)$ is the 5% damped spectral acceleration at vibration period T.

d) **Velocity Spectrum Intensity (VSI)** integrates the pseudo-spectral velocity (PSV) over a longer period range of 0.1-2.5 s (Housner, 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:

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

e) **Characteristic Intensity** I_c 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 I_c is:

$$I_c = A_{rms}^{1.5} . T_{sig}^{0.5} \tag{9}$$

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

In light of recent developments in Fourier GMMs (Bora et al., 2013, 2019; Bayless and Abrahamson, 2019), we provide a flat-file for EAS in addition to FAS from individual horizontal components. Based on the compiled and processed IM database, we next present a sensitivity analysis to verify the reliability of the processed IMs and to highlight the procedures and techniques used to refine the raw accelerograms.

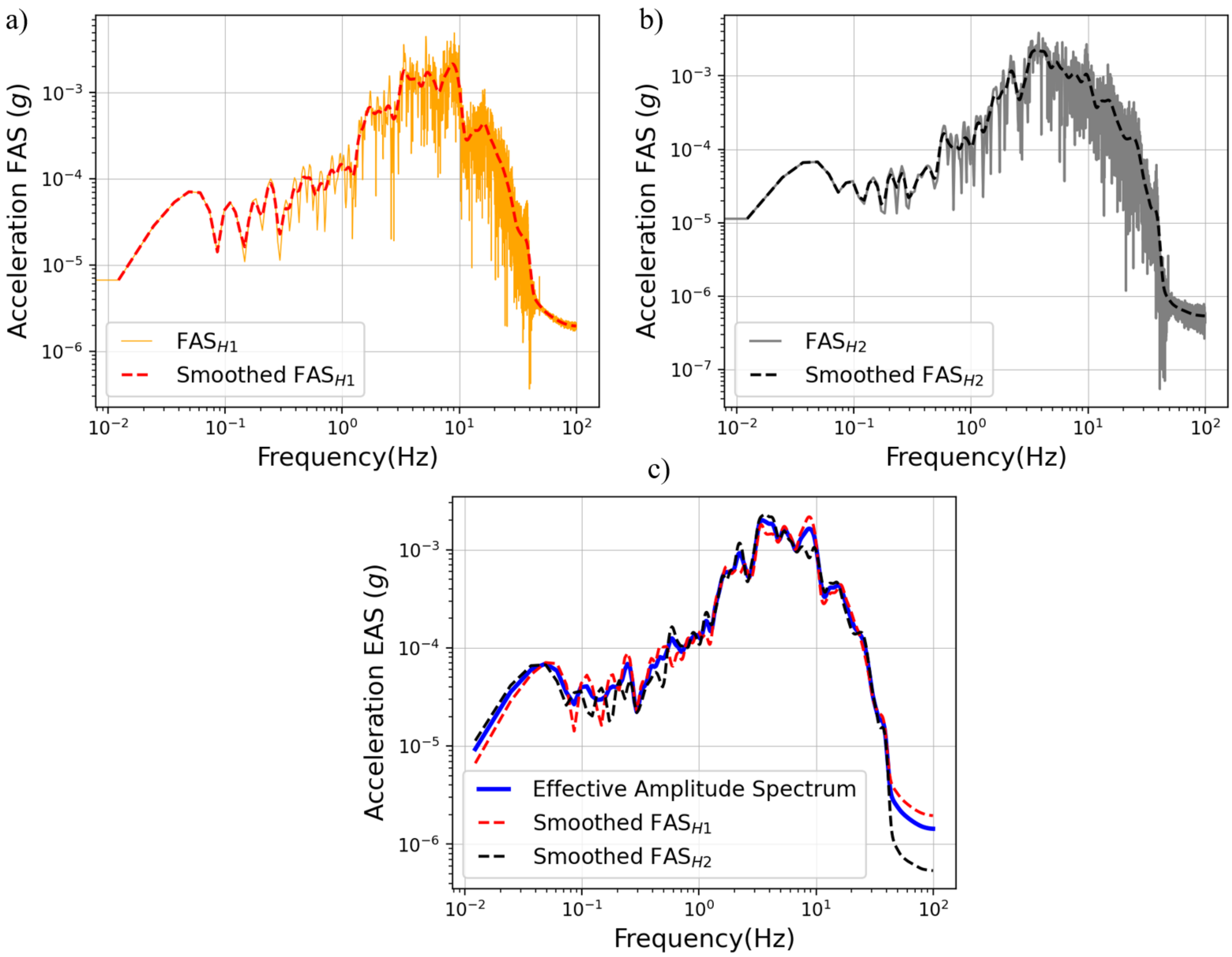

Figure 10 Variation of FAS for the (a) H1 and (b) H2 components, including smoothed curves obtained using Konno-Ohmachi smoothing (b = 40) (c) Effective amplitude plot using smoothed FAS H1 and H2 for a 2005 Chamoli earthquake record at Bageshwar station with 5.2 magnitude and 127 km epicentral distance

6 Sensitivity Analysis

In this section, we perform a sanity check on the processed database of IMs, offering valuable insights into the protocols and techniques employed to refine the raw accelerograms. To validate the integrity of the processed IMs, we performed residual analysis by comparing them with mean predictions from selected global empirical GMMs, as part of the quality control process (Manea et al., 2023). Specifically, we utilized the FAS GMM developed by Bora et al. (2019), which is calibrated using the NGA-West2 database, to evaluate the Fourier spectral ordinates of our dataset. The total residuals (Δ_{es}) for each IM were then calculated as follows:

$$\Delta_{es} = \log(y_{es}) \log(\hat{y}_{es}) \qquad (10)$$

in which y_{es} represents the observed IM for a specific event (e) and station (s) pair, similarly $\hat{y}_{es}$ denotes the median IM prediction from the GMM (in our case, Bora et al., 2019) for the same event and station combination. Predictions were computed at each frequency ordinate for 778 records within the frequency range of 0.1 to 45 Hz, aligning with the valid range of the FAS model from Bora et al. (2019). Figure 11 (a) and (b) present the total residuals for FAS and EAS respectively, for all records across a frequency range of 0.1 to 20 Hz. The mean residual is shown in red, and ± 1 standard deviation (SD) is represented by the blue lines.

The mean residual decreases with increasing frequency, indicating that the observed FAS is generally lower than the predicted values from the Bora et al. (2019) GMM, especially within the 1 to 10 Hz range. At higher frequencies, the ± 1 SD range expands, highlighting greater variability and scatter in the residuals. Note that the Bora et al. (2019) GMM is based on a subset of NG-West2 data with records predominantly from California. Clearly, there are regional differences in ground-motion and particularly in the context of Indian strong-motion data. Sharma et al. (2024) discussed that NGA-West2 models do not strictly capture the distance scaling well. Another important factor that contributes to the differences between observations and model prediction is that the Bora et al. (2019) GMM uses V_{S30} as the predictor variable to capture site-effects, in our dataset the V_{S30} values are inferred and based on surface geology. In this context, it is also worth noting

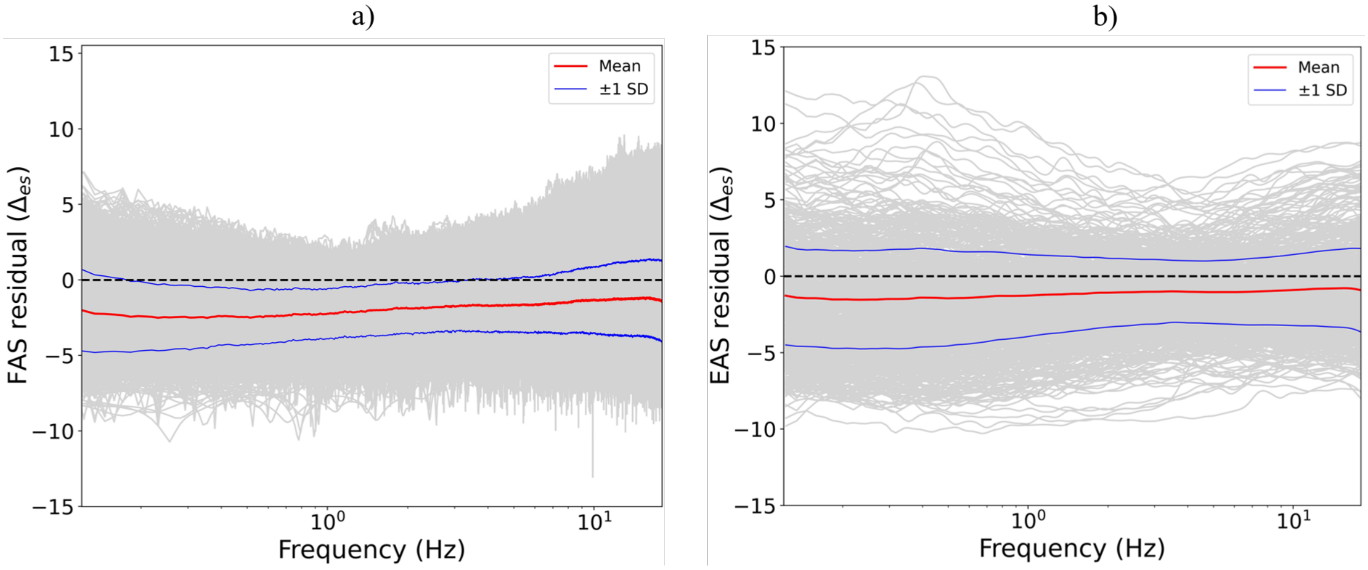

Figure 11 Variation of (a) FAS total residuals (both H1 and H2 component) (with Mean and ± 1 Standard Deviation) (b) EAS (smoothed FAS H1 and H2) total residual for 778 records from the current analysis, compared to the FAS predictions from the Bora et al. (2019) GMM.

that the majority of the records in our dataset are from small to moderate magnitude earthquakes which can result in lower average values over the entire dataset. However, as can be noticed from Figure 11, the zero-centered residual line lies within the one standard deviation bounds which provides sufficient confidence in the processed waveforms and the IMs.

7 Conclusion

This article presents detailed processing steps and methods in developing the IGM flat-file. The flat-files contain parametric tables developed from the Indian strong-motion database. The primary objective of the flat-files is to establish a standardized approach for compiling waveforms, developing flat-files, and documenting the processing protocols, with the aim of making them accessible to the scientific and engineering community in India for further applications. This is akin to similar efforts, such as NGA-West in the United States, the Engineering Strong Motion flat-files (ESM) for Europe and the Middle East. A consistent and uniform processing methodology was applied to ensure homogeneity of the strong-motion parameters, following the approach used in other regions (Puglia et al., 2018). The final flat-files include detailed metadata on seismic events and stations, various distance metrics, and IMs relevant to engineering applications. The flat-files are developed for commonly used IMs such as PGA and SA in addition to FAS and EAS separately. We additionally provide flat-files for strong ground-motion duration and integral IMs such as AI, CAV, ASI, VSI and I_c. A thorough effort is made to compile the flat-files, with a focus on organizing metadata to ensure traceability. Furthermore, the flat-files provide references for each metadata component such as events, sources, and stations. The dataset comprises approximately 778 strong-motion records from 195 earthquakes, recorded between 1986 and 2018 at 254 seismic stations, primarily associated with shallow active crustal regions in India. By incorporating seismic events with $M_w \geq 2.0$ and R_{epi} < 600 km, this dataset captures a broad range of ground motion phenomena, offering valuable insights into India's seismic activity. Importantly, it is the first of its kind study in the Indian context to compile and process the waveforms as well as the associated metadata parameters. It is possible that the metadata parameters provided in the flat-files are not free of uncertainties and assumptions. However, we have done our best to collect them from various agencies and literature. To estimate new parameters is beyond the scope of this study.

The data compiled in the flat-files can be utilized for the development of GMMs and various earthquake engineering applications, including the selection of existing GMMs for hazard assessment and adjustments of new GMMs tailored to different tectonic regimes in India. Moreover, the flat-files can aid in estimating systematic source, path, and site effects through residual decomposition, facilitating the implementation of a non-ergodic approach to seismic hazard assessment (Atik et al., 2010), especially as more data and specific details are added. Discussions within the engineering seismology community should also address the inclusion of additional metadata, such as tectonic regimes of events, stress drop, and site fundamental frequencies, depending on future research advancements. The methodology of this study can also serve as a valuable reference for uniformly processing strong-motion data in the Indian context and worldwide. Clearly, this study addresses a critical gap in India's engineering seismology domain by delivering a set of meticulously processed, metadata-rich flat-files for multiple IMs.

8 Acknowledgments

The authors gratefully acknowledge COSMOS, PESMOS and NCS for generously providing the earthquake data used in this study. We extend our sincere thanks to Dr. Govinda Niroula (University of Otago, New Zealand) for his kind assistance in accessing the Gorkha earthquake database. We are also thankful to Omkar for his valuable feedback during the development of the flat-file, which significantly improved the quality of the dataset. Finally, we express our deep appreciation to the editors and anonymous reviewers of Seismica for their insightful comments and constructive suggestions, which have greatly enhanced the clarity and impact of this manuscript.

9 Data and code availability

The earthquake flat-file produced in this study has been made available via the Figshare repository titled 'Indian Ground Motion (IGM) Database A Flat File Compilation of Strong Ground-Motion Intensity Measures' (DOI: 10.6084/m9.figshare.29530823). 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 new updates posted publicly on the portal. This ensures traceability, reproducibility, and continued usability of the dataset for researchers and practitioners. 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. All the waveform data that are being used were either requested or downloaded and are available at http://db.cosmos-eq.org, https://doi.org/10.6084/m9.figshare.19809052, https://seismo.gov.in/eventwave_datasupply and https://pesmos.org/download-ground-motion-records-2005-2017. We used PyGMT (Tian et al., 2025) to plot the figures. The scripts used for waveform reading, seismic data preprocessing, and ground-motion parameter extraction follow standard signal processing procedures and utilize commonly available libraries such as ObsPy (Beyreuther et al., 2010), NumPy, and Pandas.

10 Competing Interests

The authors declare that there are no conflicts of interest or competing interests.

References

Abrahamson, C., Shi, H., and Yang, B. Ground-motion prediction equations for Arias intensity consistent with the NGA-West2 ground-motion models. Peer report, 2016. https://peer.berkeley.edu/sites/default/files/webpeer-2016-05_charlotte_abrahamson_hao-jun_michael_shi_brian_yang.pdf.

Afshari, K. and Stewart, J. P. Physically Parameterized Prediction Equations for Significant Duration in Active Crustal Regions. *Earthquake Spectra*, 32(4):2057–2081, Nov. 2016. doi: 10.1193/063015eqs106m.

Akkar, S. and Cagnan, Z. A Local Ground-Motion Predictive Model for Turkey, and Its Comparison with Other Regional and Global Ground-Motion Models. *Bulletin of the Seismological Society of America*, 100(6):2978–2995, Dec. 2010. doi: 10.1785/0120090367.

Akkar, S., Sandıkkaya, M. A., and Bommer, J. J. Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bulletin of Earthquake Engineering*, 12(1):359–387, May 2013. doi: 10.1007/s10518-013-9461-4.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L. NGA-West2 Database. *Earthquake Spectra*, 30(3):989–1005, Aug. 2014. doi: 10.1193/070913eqs197m.

Arango, M. C., Strasser, F. O., Bommer, J. J., Boroschek, R., Comte, D., and Tavera, H. A strong-motion database from the Peru–Chile subduction zone. *Journal of Seismology*, 15(1): 19–41, Aug. 2010. doi: 10.1007/s10950-010-9203-x.

Arias, A. *A measure of earthquake intensity. Seismic Design for Nuclear Power Plants*. The MIT Press, Cambridge, Massachusetts, U.S.A, 1970.

Atik, L. A., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N. The Variability of Ground-Motion Prediction Models and Its Components. *Seismological Research Letters*, 81 (5):794–801, Aug. 2010. doi: 10.1785/gssrl.81.5.794.

Bansal, B. K., Pandey, A. P., Singh, A. P., Suresh, G., Singh, R. K., and Gautam, J. L. National Seismological Network in India for Real-Time Earthquake Monitoring. *Seismological Research Letters*, 92(4):2255–2269, Mar. 2021. doi: 10.1785/0220200327.

Bastías, N. and Montalva, G. A. Chile Strong Ground Motion Flatfile. *Earthquake Spectra*, 32(4):2549–2566, Nov. 2016. doi: 10.1193/102715eqs158dp.

Bayless, J. and Abrahamson, N. A. Summary of the BA18 Ground-Motion Model for Fourier Amplitude Spectra for Crustal Earthquakes in California. *Bulletin of the Seismological Society of America*, 109(5):2088–2105, Sept. 2019. doi: 10.1785/0120190077.

Benjamin, J. R. A criterion for determining exceedance of the operating basis earthquake. In *EPRI NP-5930*. 1988.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, 81(3):530–533, May 2010. doi: 10.1785/gssrl.81.3.530.

Bhandari, L., Fuloria, R., and Sastri, V. Stratigraphy of Assam Valley, India. *AAPG Bulletin*, 57(4):642–654, 1973.

Bilham, R. Himalayan earthquakes: a review of historical seismicity and early 21st century slip potential. *Geological Society, London, Special Publications*, 483(1):423–482, Feb. 2019. doi: 10.1144/sp483.16.

Bindi, D., Kotha, S.-R., Weatherill, G., Lanzano, G., Luzi, L., and Cotton, F. The pan-European engineering strong motion (ESM) flatfile: consistency check via residual analysis. *Bulletin of Earthquake Engineering*, 17(2):583–602, Sept. 2018. doi: 10.1007/s10518-018-0466-x.

Bommer, J. and Martínez-Pereira, A. Strong-motion parameters: definition, usefulness and predictability. In *Proc. of the 12th World Conference on Earthquake Engineering*, Auckland, New Zealand, 2000.

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

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

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

Boore, D. M. and Akkar, S. Effect of causal and acausal filters on elastic and inelastic response spectra. *Earthquake Engineering & Structural Dynamics*, 32(11):1729–1748, June 2003. doi: 10.1002/eqe.299.

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

Bora, S. S., Scherbaum, F., Kuehn, N., and Stafford, P. Fourier spectral- and duration models for the generation of response spectra adjustable to different source-, propagation-, and site conditions. *Bulletin of Earthquake Engineering*, 12(1):467–493, July 2013. doi: 10.1007/s10518-013-9482-z.

Bora, S. S., Cotton, F., and Scherbaum, F. NGA-West2 Empirical Fourier and Duration Models to Generate Adjustable Response Spectra. *Earthquake Spectra*, 35(1):61–93, Feb. 2019. doi: 10.1193/110317eqs228m.

Bozorgnia, Y. and Campbell, K. The vertical-to-horizontal response spectral ratio and tentative procedures for developing simplified V/H and vertical design spectra. *Journal of Earthquake Engineering*, 8(2):175–207, Mar. 2004. doi: 10.1080/13632460409350486.

Campbell, K. and Bozorgnia, Y. Next generation attenuation (NGA) empirical ground motion models: can they be used in Europe. In *Proceedings, First European Conference on Earthquake Engineering and Seismology*, 2006.

Campbell, K. W. and Bozorgnia, Y. A Ground Motion Prediction Equation for the Horizontal Component of Cumulative Absolute Velocity (CAV) Based on the PEER-NGA Strong Motion Database. *Earthquake Spectra*, 26(3):635–650, Aug. 2010. doi: 10.1193/1.3457158.

Campbell, K. W. and Bozorgnia, Y. Ground Motion Models for the Horizontal Components of Arias Intensity (AI) and Cumulative Absolute Velocity (CAV) Using the NGA-West2 Database. *Earthquake Spectra*, 35(3):1289–1310, Aug. 2019. doi: 10.1193/090818eqs212m.

Chiou, B. J. and Youngs, R. R. An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra*, 24(1):173–215, Feb. 2008. doi: 10.1193/1.2894832.

Chopra, A. K. and Chakrabarti, P. The Koyna Earthquake: A review of the seismicity and strong motion records. *Bulletin of the Seismological Society of America*, 63(2):381–397, Apr. 1973. doi: 10.1785/bssa0630020381.

Converse, A. and Brady, A. G. *BAP basic strong-motion accelerogram processing software version 1.0*. 1992. doi: 10.3133/ofr92296a.

Dai, J.-C., Wang, D.-S., Chen, X.-Y., Zhang, R., and Sun, Z.-G. Evaluation of ground motion intensity measures for time-history dynamic analysis of isolated bridges. *Structures*, 55:1306–1319, Sept. 2023. doi: 10.1016/j.istruc.2023.06.007.

Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E. A Flatfile for the KiK-net Database Processed Using an Automated Protocol. *Earthquake Spectra*, 32(2): 1281–1302, May 2016. doi: 10.1193/071214eqs106.

Dobry, R., Idriss, I., and Ng, E. Duration characteristics of horizontal components of strong-motion earthquake records. *Bulletin of the Seismological Society of America*, 68(5):1487–1520, 1978. doi: 10.1785/BSSA0680051487.

Douglas, J. What is a Poor Quality Strong-Motion Record? *Bulletin of Earthquake Engineering*, 1(1):141–156, Jan. 2003. doi: 10.1023/a:1024861528201.

Douglas, J. and Boore, D. M. High-frequency filtering of strong-motion records. *Bulletin of Earthquake Engineering*, 9(2): 395–409, Sept. 2010. doi: 10.1007/s10518-010-9208-4.

Farhadi, A. and Pezeshk, S. A Referenced Empirical Ground-Motion Model for Arias Intensity and Cumulative Absolute Velocity Based on the NGA-East Database. *Bulletin of the Seismological Society of America*, 110(2):508–518, Mar. 2020. doi: 10.1785/0120190267.

Goulet, C., Kottke, A., Boore, D., Bozorgnia, Y., Hollenback, J., Kishida, T., Der Kiureghian, A., Ktenidou, O., Kuehn, N., Rathje, E., et al. Effective amplitude spectrum (EAS) as a metric for ground motion modeling using Fourier amplitudes. In *2018 Seismological Society of America Annual Meeting*, Miami, Florida, 2018.

Housner, G. Behavior of Structures During Earthquakes. *Journal of the Engineering Mechanics Division*, 85(4):109–129, Oct. 1959. doi: 10.1061/jmcea3.0000102.

Jain, S. K. Earthquake safety in India: achievements, challenges and opportunities. *Bulletin of Earthquake Engineering*, 14(5): 1337–1436, Apr. 2016. doi: 10.1007/s10518-016-9870-2.

Jain, S. K. and Agrawal, P. Earthquake engineering capacity building in educational sector in India. In *Proceedings of the 13th world conference on earthquake engineering. Vancouver, Canada*, 2004.

Kalkan, E. An AutomaticP-Phase Arrival-Time Picker. *Bulletin of the Seismological Society of America*, 106(3):971–986, May 2016. doi: 10.1785/0120150111.

Kamiyama, M. Effects of subsoil conditions and other factors on the duration of earthquake ground shaking. In *Proceedings, 8th World Conference on Earthquake Engineering, San Francisco*, volume 2, page 793–800, 1984.

Kapoor, H. and Singh, G. Extra-peninsular Gondwana basins - stratigraphy and evolution. *Journal of Palaeosciences*, 36: 312–325, Dec. 1987. doi: 10.54991/jop.1987.1590.

Kempton, J. J. and Stewart, J. P. Prediction Equations for Significant Duration of Earthquake Ground Motions considering Site and Near-Source Effects. *Earthquake Spectra*, 22(4):985–1013, Nov. 2006. doi: 10.1193/1.2358175.

Konno, K. and Ohmachi, T. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *Bulletin of the Seismological Society of America*, 88(1):228–241, Feb. 1998. doi: 10.1785/bssa0880010228.

Kumar, A., Mittal, H., Sachdeva, R., and Kumar, A. Indian Strong Motion Instrumentation Network. *Seismological Research Letters*, 83(1):59–66, Jan. 2012. doi: 10.1785/gssrl.83.1.59.

Lanzano, G., Sgobba, S., Luzi, L., Puglia, R., Pacor, F., Felicetta, C., D'Amico, M., Cotton, F., and Bindi, D. The pan-European Engineering Strong Motion (ESM) flatfile: compilation criteria and data statistics. *Bulletin of Earthquake Engineering*, 17(2): 561–582, Sept. 2018. doi: 10.1007/s10518-018-0480-z.

Liao, B.-Y., Huang, H.-C., and Xie, S. 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, Jan. 2022. doi: 10.3390/app12010494.

Luzi, L., Puglia, R., Russo, E., D'Amico, M., Felicetta, C., Pacor, F., Lanzano, G., Çeken, U., Clinton, J., Costa, G., Duni, L., Farzanegan, E., Gueguen, P., Ionescu, C., Kalogeras, I., Özener, H., Pesaresi, D., Sleeman, R., Strollo, A., and Zare, M. The Engineering Strong-Motion Database: A Platform to Access Pan-European Accelerometric Data. *Seismological Research Letters*, 87(4): 987–997, June 2016. doi: 10.1785/0220150278.

Mai, P. M. and Thingbaijam, K. SRCMOD: An online database of finite-fault rupture models. *Seismological Research Letters*, 85 (6):1348–1357, 2014.

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

Meimandi-Parizi, A., Daryoushi, M., Mahdavian, A., and Saffari, H. Ground-Motion Models for the Prediction of Significant Duration Using Strong-Motion Data from Iran. *Bulletin of the Seismological Society of America*, 110(1):319–330, Jan. 2020. doi: 10.1785/0120190109.

Mittal, H., Gupta, S., Srivastava, A., Dubey, R., and Kumar, A. National strong motion instrumentation project: an overview. In *13th symposium on earthquake engineering, Indian Institute of Technology, Roorkee*, pages 18–20, 2006.

Pacor, F., Paolucci, R., Ameri, G., Massa, M., and Puglia, R. Italian strong motion records in ITACA: overview and record processing. *Bulletin of Earthquake Engineering*, 9(6):1741–1759, June 2011. doi: 10.1007/s10518-011-9295-x.

Park, Y. and Ang, A. H. Mechanistic Seismic Damage Model for Reinforced Concrete. *Journal of Structural Engineering*, 111(4):722–739, Apr. 1985. doi: 10.1061/(asce)0733-9445(1985)111:4(722).

Puglia, R., Russo, E., Luzi, L., D'Amico, M., Felicetta, C., Pacor, F., and Lanzano, G. Strong-motion processing service: a tool to access and analyse earthquakes strong-motion waveforms. *Bulletin of Earthquake Engineering*, 16(7):2641–2651, Jan. 2018. doi: 10.1007/s10518-017-0299-z.

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

Rao, A. and Varma, C. Strong motion instrumentation in India. *Proc. Indian Acad. Sci. (Earth Planet. Sci*, 90(1):127–134, 1981.

Sali, S. Some aspects of the Pleistocene stratigraphy of Peninsular India. *Bulletin of the Deccan College Research Institute*, 31(1/2): 70–80, 1970.

Scordilis, E. M. Empirical global relations converting MS and mb to moment magnitude. *Journal of Seismology*, 10(2):225–236, Apr. 2006. doi: 10.1007/s10950-006-9012-4.

Sharma, S., Mannu, U., and Singh Bora, S. Epistemic Uncertainty in Ground-Motion Characterization in the Indian Context: Evaluation of Ground-Motion Models (GMMs) for the Himalayan Region. *Seismological Research Letters*, 95(3):1718–1734, Dec. 2024. doi: 10.1785/0220230157.

Shigefuji, M., Takai, N., Bijukchhen, S., Ichiyanagi, M., Rajaure, S., Dhital, M. R., Paudel, L. P., and Sasatani, T. Strong ground motion data of the 2015 Gorkha Nepal earthquake sequence in the Kathmandu Valley. *Scientific Data*, 9(1), Aug. 2022. doi: 10.1038/s41597-022-01634-6.

Sriwastav, R., Yedulla, J., and Raghukanth, S. A non-parametric model of ground motion parameters for shallow crustal earthquakes in Europe. *Soil Dynamics and Earthquake Engineering*, 186:108923, 2024. 10.1016/j.soildyn.2024.108923.

Stafford, P. J., Berrill, J. B., and Pettinga, J. R. New predictive equations for Arias intensity from crustal earthquakes in New Zealand. *Journal of Seismology*, 13(1):31–52, July 2008. doi: 10.1007/s10950-008-9114-2.

Takai, N., Shigefuji, M., Rajaure, S., Bijukchhen, S., Ichiyanagi, M., Dhital, M. R., and Sasatani, T. Strong ground motion in the Kathmandu Valley during the 2015 Gorkha, Nepal, earthquake. *Earth, Planets and Space*, 68(1), Jan. 2016. doi: 10.1186/s40623-016-0383-7.

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

Tian, D., Leong, W. J., Fröhlich, Y., Grund, M., Schlitzer, W., Jones, M., Toney, L., Yao, J., Tong, J.-H., Magen, Y., Materna, K., Belem, A., Newton, T., Anant, A., Ziebarth, M., Quinn, J., Uieda, L., and Wessel, P. PyGMT: A Python interface for the Generic Mapping Tools, Oct. 2025. doi: 10.5281/zenodo.17156962.

Trifunac, M. and Brady, A. A study on the duration of strong earthquake ground motion. *Bulletin of the Seismological Society of America*, 65(3):581–626, 1975. doi: 10.1785/BSSA0650030581.

Von Thun, J. Earthquake ground motions for design and analysis of dams, 1988.

IGM_Flat-file Header	General Description
Record_Id	Recording ID (separate for each record)
Event_Id	Earthquake ID (separate for each earthquake event)
Eqk_Name	Earthquake Name (full name of the earthquake as provided by the different networks)
Year	Year in which earthquake occurred
MOD	Month and date (of the earthquake occurrence)
HRMN	Hours and minutes (Time of earthquake)
Time_Zones	IST - Indian Standard Time (UTC+05:30); UTC - Coordinated Universal Time
Stn_Id	Station ID
Stn_Name	Station name
Stn_code	Station code
Surface Geology	Reported as provided by the respective networks or compiled from published geological/literature sources; classifications follow either network-specific schemes or NEHRP categories where available.
Region	Place where earthquake has occurred
Network	The agency that provided the data
Eqk_Lat	Earthquake Latitude
Eqk_Long	Earthquake Longitude
Stn_Lat	Station Latitude
Stn_Long	Station Longitude
Samp_Freq(Hz)	Sampling Frequency (in Hertz)
Low_Cut	High-pass filter cutoff frequency (in Hertz)
High_Cut	Low-pass filter cutoff frequency (in Hertz)
Network_Site_class	Site class provided by the different networks (when available)
NEHRP_Site_class	NEHRP site classification
V_{S30} (m/s)	Shear wave velocity provided by different networks
Given_Mag	Magnitude value (same as provided by networks)
Moment_mag	Moment magnitude
Depth(km)	Depth of an earthquake (in km)
Hypo_Dist(km)	Hypocentral distance (in km)
Epi_Dist(km)	Epicentral distance (in km)
Given_PGA(g)	Peak Ground Acceleration provided by the networks (in g)
PGA_GeoMean(g)	Peak Ground Acceleration, calculated using the geometric mean of the two horizontal components (in g)
PGV_GeoMean(m/s)	Peak Ground Velocity, computed by integrating two horizontal components for velocity time histories using the trapezoidal rule, then calculating geometric mean (in m/s)
PGD_GeoMean(m)	Peak Ground Displacement, computed by double integrating of two horizontal components, calculating the geometric mean of the resulting displacements (in m)
ASI(cm/s)	Acceleration Spectrum Intensity (ASI)
VSI(cm)	Velocity Spectrum Intensity (VSI)
$I_c(cm^3/s^{5/2})$	Characteristic Intensity (I_c)
T_(0.01 - 10)	Spectral Acceleration for the period range of 0.01 to 10s (in g)[Refer IGM sheet1]
FAS(f_0.01 - f_25.0)(separately for H1 and H2 component)	Fourier amplitude spectrum (in g) for different frequency (in Hz) [Refer IGM sheet 2 [H1 component] and sheet 3 [H2 component]]
AI (t_0.01 - t_80)	Arias Intensity in m/s for time between 0 - 80s [Refer IGM sheet 4]
CAV (t_0.01 - t_80)	Cumulative absolute velocity in m/s for time between 0 - 80s [Refer IGM sheet 5]
D5-75, D20-80, D5-95	Significant duration in sec [Refer IGM sheet 6]
Rotation independent measures of IMs	RotD0 [sheet 7], RotD50 [sheet 8] and RotD100 [sheet 9] over the period range of 0.1 to 10s
EAS (f_0.01 - f_25.0)	Effective Amplitude Spectrum (in g) for different frequency (in Hz) [Refer IGM sheet 10]

Table 1 Key Seismological Parameters in the Flat-file: Acronyms, Headers, and Descriptions