The recently compiled strong motion databank of Iran

Sahar Shokouhirad
IIEES: International Institute of Earthquake Engineering and Seismology
Anooshiravan Ansari
a.ansari@iiees.ac.ir
IIEES: International Institute of Earthquake Engineering and Seismology https://orcid.org/0000-0003-1241-8554
Mohsen Ghafory-Ashtiany
IIEES: International Institute of Earthquake Engineering and Seismology

Research Article

Keywords: Iran, Integrated catalog, Site classification, Strong ground motion records, Ground motion data processing

Posted Date: December 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3768685/v1

Version of Record: A version of this preprint was published at Bulletin of Earthquake Engineering on December 3rd, 2024. See the published version at https://doi.org/10.1007/s10518-024-02052-2.

The recently compiled strong motion databank of Iran

Sahar Shokouhirad [1]. Anooshiravan Ansari [2]. Mohsen Ghafory-Ashtiany [3].

Abstract

This paper presents a comprehensive and integrated databank of the Iranian strong ground motions that occurred from 1973 to 2018. The databank consists of 7196 three-component acceleration records from 3180 earthquakes and 1157 stations in Iran. In this paper, the characteristics of the Iranian strong ground motion data are presented in terms of event, station, and recording distributions. The events are characterized by magnitude in the range 2.4-7.7. Shear wave velocity has been measured and reported at 603 strong motion stations of the databank. In this study, three different empirical techniques are applied to classify the stations. A new method is proposed for site classification based on the correlation coefficient between the horizontal-to-vertical (H/V) response spectral ratios of the ground motion records recorded by each station. It is noticeable that the raw accelerograms have been uniformly processed in the entire databank using the filtering and wavelet de-noising methods to remove high- and low-frequency noise. Moreover, by comparison between the Fourier Amplitude Spectrum (FAS) of the noises detected in all acceleration and velocity time series by the filtering and the wavelet de-noising methods, it was determined that the mean and mode of FAS of the noises detected by the wavelet de-noising method in most of the frequencies is higher than mean and mode of FAS of the noises detected by the filtering method.

Keywords Iran · Integrated catalog · Site classification · Strong ground motion records · Ground motion data processing

✉ Anooshiravan Ansari
a.ansari@iiees.ac.ir

[1] PhD Candidate at International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran

[2] Associate Professor at International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran, https://orcid.org/0000-0003-1241-8554

[3] Distinguished Professor at International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran, https://orcid.org/0000-0002-3821-9193

1 Introduction

In general, the progress of earthquake engineering knowledge is due to the analysis of accelerograms recorded during different earthquakes. Strong motion records play also an important role in development of ground motion prediction equations (GMPE) (Ghasemi et al. 2009a; Hamzehloo and Mahood 2012; Sedaghati and Pezeshk 2017; Zafarani et al. 2018; Farajpour et al. 2019) which are widely used in seismic hazard studies. These records are also an important basis for development of vulnerability curves and fatality models in risk and loss analyses (Firuzi et al. 2020).

The Iranian plateau is a part of the Alpine–Himalayan orogenic belt, which has a high level of seismic activity. This region has experienced several destructive earthquakes including 1978 Tabas earthquake (M_w7.3), 1990 Rudbar-Manjil earthquake (M_w 7.4), 2003 Bam earthquake (M_w 6.6), 2012 Varzagan earthquake (M_w 6.5), and 2017 Ezgeleh earthquake (M_w 7.4).

In Iran, the first strong motion instruments were installed in 1973. This marks the establishment of a nationwide strong motion network that has been operated by the Iranian Strong Motion Network (ISMN) of the Road, Housing and Urban Development Research Center (BHRC). Instruments first used were of Kinematics SMA-1 type (analog recorders), that recorded the first mainshock at Jahrom station during the M_w 5.2, February 24, 1973, earthquake. Between 1973 and 1992, 276 units of SMA-1 analog accelerographs were installed. From 1994, digital accelerographs started to be used in Iran. Since then, the number of strong motion stations has been gradually increased. Most of the accelerograph stations are installed in seismically active or in densely populated and industrialized areas. By the end of 2020, the ISMN database includes more than 15,600 three-component time histories, with magnitudes ranging from 2.0 to 7.7 and rupture distances ranging from a few kilometers to 1000 km, recorded at about 1518 active stations. Major earthquake events are accessible on the BHRC website. (https://smd.bhrc.ac.ir/Portal/en/Search/BigQuakes)

Since Iran has one of the richest and oldest strong motion datasets in the world, development of a strong ground motion dataset with appropriate quality and content would be an essential component in seismic hazard and risk studies and for engineering purposes. This database was used by several researchers to develop GMPEs for Iranian Plateau (e.g. Ghasemi et al. 2009a), Hamzehloo and Mahood (2012), Sedaghati and Pezeshk (2017), Zafarani et al. (2018) and Farajpour et al. (2019)). This dataset was also used to develop vulnerability and fatality models of Iran (Firuzi et al. 2020, 2022). Zafarani and Soghrat (2017) published a comprehensive databank of ground motions to the end of 2013, which consists of 2286 three-component records from 461 Iranian earthquakes with moment magnitudes from 3.9 to 7.3. Farajpour et al. (2018) developed a comprehensive databank of shallow earthquake ground motions to the end of 2013 with depths less than 35 km, which consists of 860 three-component records from 183 Iranian earthquakes with moment magnitudes from 5.0 to 7.4 and rupture distances less than 120 km.

This paper aims to develop a new catalog with comprehensive, reliable, and unified metadata from Iran for earthquake, structural, and geotechnical engineers as well as seismologists for structural design, nonlinear dynamic analysis, and hazard and risk analyses. In this regard, 7196 accelerograms of 3180 earthquakes recorded by 1157 stations from 1973 to 2018 with moment magnitudes ranging from 2.4-7.7 were analyzed in this study. The main features of the earthquakes, stations and accelerograms such as location, magnitude, focal depth, style and main plane of faulting, parameters for site characterization, station coordinates, several distance metrics, unprocessed and processed acceleration time histories, corner frequencies and power spectral density of the noises detected are presented in the dataset.

For 603 out of 1157 strong motion stations, the value of shear wave velocity over 30 m (V_{s30}) from all the stations were determined by BHRC (https://smd.bhrc.ac.ir/Portal/en/Search/Stations). To provide an estimation of V_{s30} for other stations, a set of empirical techniques is used to classify strong motion sites stations. An effort was done to provide a more satisfying model for prediction of site class of different stations. The reliability of the classification schemes is carried out using the same dataset used to define the average horizontal-to-vertical response spectral ratios (H/V), and the results of different schemes are represented and compared. Finally, all stations without measurement of V_{s30} in the databank have been classified using the proposed empirical techniques.

A special attention was also paid to the effect of record correction on the final acceleration records. In this regard, the results of conventional filtering method are compared to the wavelet de-noising approach (Ansari et al. 2007, 2010) to determine which method could better identify noise of the records. The results indicate that wavelet de-noising has more capability to identify noise in whole frequency range of interest.

In the following, first, the general overview of strong motion records of Iran is presented. The process of compilation of databank is the construction a relational database in PostgreSQL. Such a database could facilitate future applications of the databank, decrease data inconsistencies and also provide a basis for developing different software applications. Although description details of such a relational database are beyond the scope of this paper, however, the high-level structure of the database developed in the International Institute of Earthquake Engineering and Seismology (IIEES) is described in this paper. It is worth noting that the term databank is a generic term meaning any collection of data in any form and database is a technical term referring to a collection of data managed by a software called a Database Management System (DBMS).

In this paper, a general overview of Iran acceleration records, the introduction of various references of databank development, main parts of the databank, compiling of the main features and related information of the earthquakes, stations, and ground motion records in the metadata, using the correlation coefficient parameter for quantitative site classification, the data processing, and the estimation of the various strong motion intensity measures are discussed.

2 General overview of Iran ground motion records

Figure 1 shows the number of records in the dataset as a function of time for different distance ranges (i.e. epicentral distances less than 50 and 10 km) and the location of the stations belonging to the ISMN (1518 stations). The amount of data is nearly constant until 1993 and then it significantly increases as a consequence of the rapid growth of the number of strong motion stations. In other words, since 1994, the number of records and earthquakes considerably increased because of the rapid development of the number of digital strong motion stations. Specific seismic activity close to the stations or improvement in the operational status of the instruments could be considered as other causes of this growth. To elucidate the effect of such growth, it is worth mentioning that the 2004 Baladeh earthquake with M_w 6.2 in the Alborz–Azarbaijan region was recorded by about 148 stations, while the greater Rudbar earthquake in 1990 with M_w 7.3 was recorded only by 23 stations.

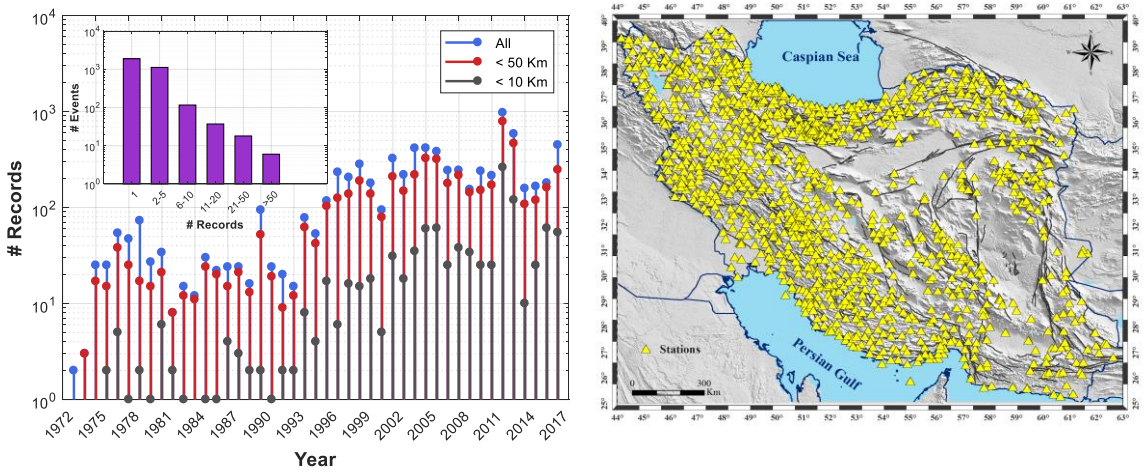


Fig. 1 Distribution of the number of ground motion records versus years since 1973 for different distances ranges. In the box, histogram of the number of ground motion records per events (*left*) and map of Iranian Strong Motion Network (ISMN) (*right*)

According to Figure 1, more than 70% of the strong motion records have epicentral distance less than 50 km, but only less than 15% of these time histories records the motion in distances less than 10 km. Regarding number of stations that record each earthquake, about 60 percent of events were recorded by only one station. As it is clear from this figure, there is a jump in the number of records from mid 90s, mainly because of the development of the network.

Figure 2 shows the distribution of peak ground acceleration (*PGA*) and velocity (*PGV*) values for the databank in the period 1973 – 2018. A total of 1391 waveforms (about 19% of the total ground motion records) have *PGA* > 50 cm/s^2 while 432 recordings (about 6% of the total records) have *PGV* > 5 cm/s. In both cases, the maximum of two horizontal components was considered.

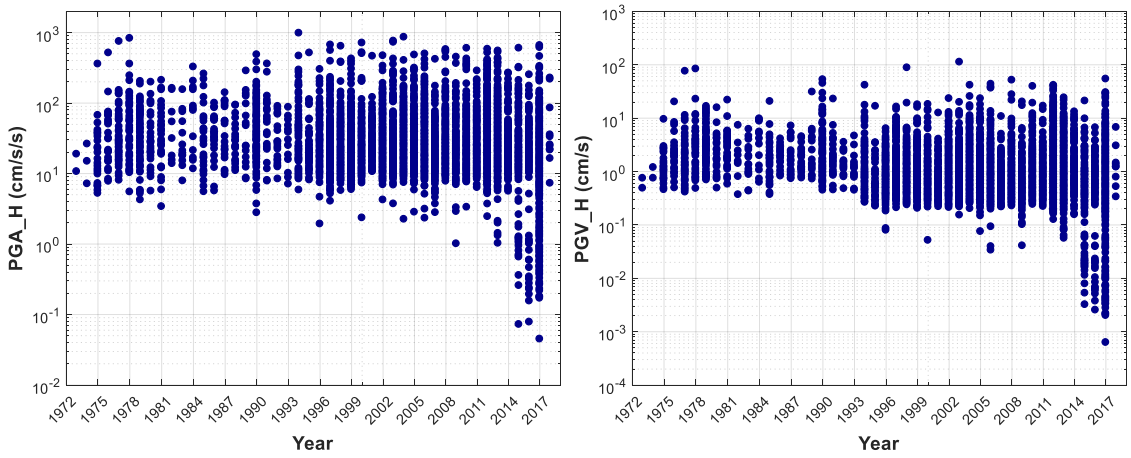
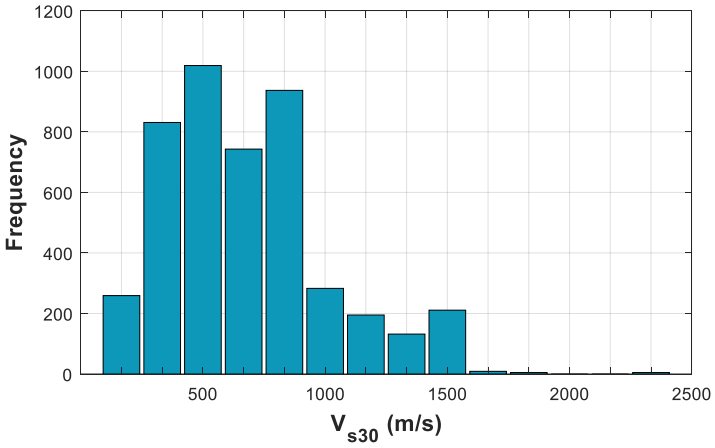


Fig. 2 Distribution with time of maximum horizontal *PGA* (*left*) and *PGV* (*right*) of Iranian ground motion records

The histogram of records versus V_{s30} is shown in Fig. 3. Among 1157 strong motion stations, in 603 cases the V_{s30} is measured and reported by BHRC (https://smd.bhrc.ac.ir/Portal/en/Search/Stations). Moreover, only 14 accelerograms are recorded on the very soft soil (V_{s30} less than 175 m/s) based on the Iranian national seismic standard. Moreover, 44% of acceleration time series are recorded on the soils with the V_{s30} between 175 and 750 m/s, and the rest of data corresponds to the hard rock soil condition.

Fig. 3 Frequency of the average shear wave velocity over 30 m (V_{s30}) of the Iranian ground motion records databank



3 Databank development

The established databank is organized into four main sections: (1) characteristics of earthquakes, (2) characteristics of strong motion stations, (3) characteristics of ground motion records, and (4) engineering parameters. Figure 4 describes the main sections of the developed databank along with their related parameters, national and international seismic agencies for gathering information, and database structure. Further information regarding the evaluation database is available in the electronic supplement to this article.

4 Earthquakes

The earthquake information gathered for each event in the databank consists of the origin time, epicentral coordinates, earthquake magnitude in various scales (moment magnitude (M_w), surface-wave magnitude (M_s), body-wave magnitude (m_b), local magnitude (M_L) and Nuttli magnitude (M_N)), focal depth, main plane and style-of-faulting (SoF).

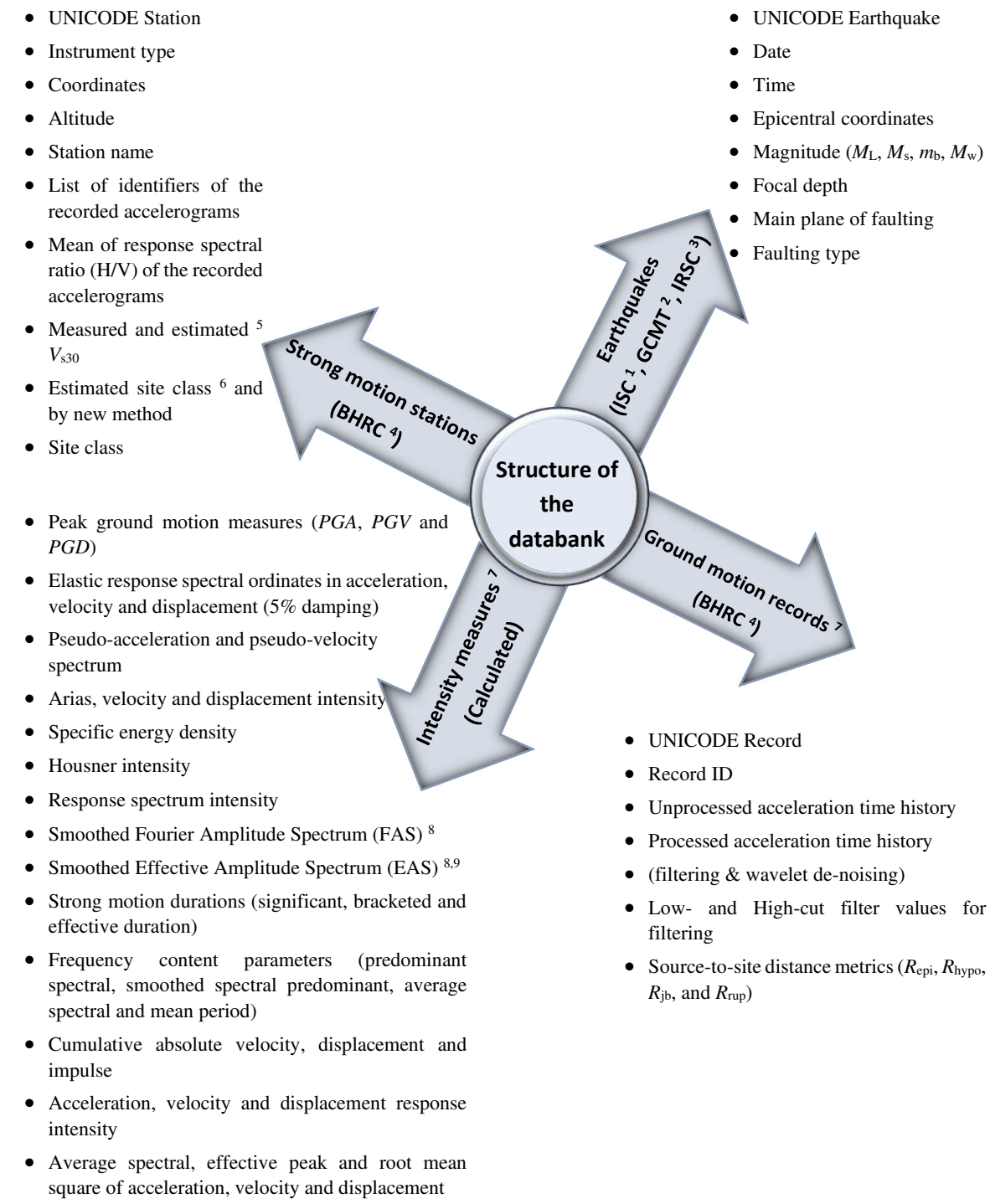


Fig. 4 Schematic structure of the accelerogram databank contents

[1] International Seismological Center (ISC) http://www.isc.ac.uk, [2] Harvard Global Centroid Moment Tensor (GCMT) catalog http://www.globalcmt.org, [3] The Iranian Seismological Centre, Institute of Geophysics, University of Tehran (IRSC) http://www.irsc.ut.ac.ir, [4] Road, Housing and Urban Development Research Center (BHRC) http://www.bhrc.ac.ir, [5] Based on the method proposed by Wald and Allen (2007), [6] Using H/V method based on the proposed method of Ghasemi et al. (2009b), [7] Characteristics of records and engineering parameters are provided for each record component (2 horizontal and 1 vertical), [8] Smoothed using the Konno and Ohmachi (1998) function with the coefficient of bandwidth smoothing window in $\log_{10}$ units b_w = 1/50 in the frequency range 0.001-50 Hz, [9] Effective Amplitude Spectrum (EAS) as defined by Pacific Earthquake Engineering Research Center, PEER, (Goulet et al. 2018), is provided in the frequency range 0.001-50 Hz.

The attribution of events locations (geographical coordinates and depth) and magnitudes in this databank is obtained according to the Bulletin of International Seismological Center (ISC) as a main and primary reference, which collects the data from over 130 agencies worldwide and provides them online soon after being received, the policy for databank compilation is to use only the ISC reviewed data, available after 24 months (Willemann and Storchak 2001; Storchak et al. 2020). As a secondary reference, the national catalog of the Iranian Seismological Center (IRSC) has been considered to collect event information.

The earthquake epicenter location (geographical coordinates) error of the IRSC catalog from the reviewed ISC bulletin and the comparison between the focal depth of earthquakes in the databank obtained from these two international and national bulletins are shown in figures 5a and 5b, respectively. According to Fig. 5a, the coordinates of the epicenter announced by the two sources have a greater difference in smaller moment magnitudes (2 to 5), and with the increase in magnitude, this difference decreases, so that the difference reaches below 25 km in the moment magnitude between 6 to 8. Obvious differences are observed from focal depth, especially for the shallow crustal earthquakes. In general, the focal depth of the event is a location parameter often poorly constrained, and thus significant differences can be observed among different authoritative sources (Naserieh et al. 2019).

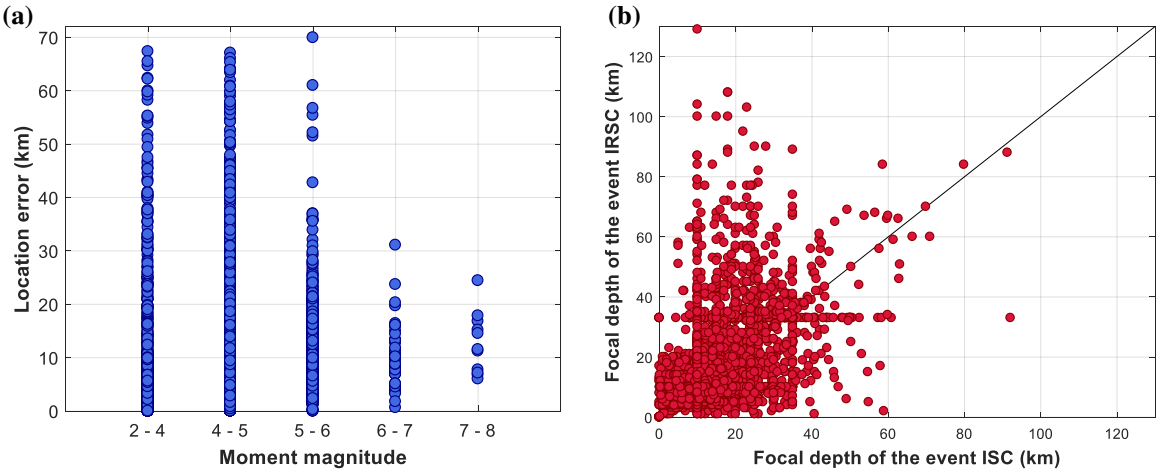


Fig. 5 Comparison between databank information obtained from ISC and IRSC references in terms of **a** epicenter location error versus moment magnitude and **b** focal depth

In the databank, moment magnitude is provided from international and national bulletins for only 12% of events (387 of 3180 events), for the remaining events we used the magnitude conversion relationships of Mousavi-Bafrouei and Mahani (2020) to obtain the more homogenous magnitude information and increase the number of earthquakes associated with moment magnitude values. In Fig. 6, the histogram of the reported moment magnitude compared to the converted moment magnitude of the earthquakes is shown in terms of magnitude. As it is clear from this figure, in smaller magnitudes ($M_w \leq 6$), the number of events with converted magnitudes is much higher than the number of events with reported magnitudes. Moreover, in higher magnitudes ($M_w > 6$), the moment magnitude has been reported in all events.

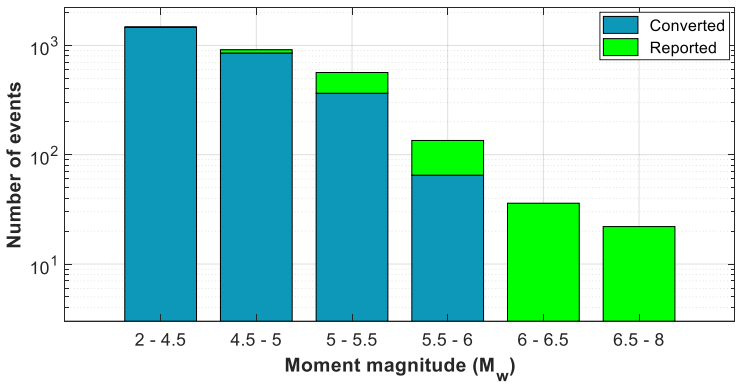


Fig. 6 Distribution of the events as a function of moment magnitude in terms of the reported and converted M_w

There are different definitions such as Frohlich and Apperson (1992), Zoback (1992), Atkinson and Boore (2007), etc. to the determination of the focal mechanisms and for the classification of style-of-faulting (SoF). In this study, earthquakes were classified by SoF using the criteria of Frohlich and Apperson (1992) based on the plunges of the eigenvectors of the moment tensor obtained from the Harvard Global Centroid Moment Tensor (GCMT) catalog. The method of Frohlich and Apperson uses the plunge angles of the *P*, *T*, and *B* axes for the fault plane solutions, they define earthquakes as normal, strike-slip, or reverse if the dip of the *P* axes exceeds 60°, if the dip of the *B* axes exceeds 60° and if the dip of the *T* axis exceeds 50°, respectively. Odd earthquakes are all other events.

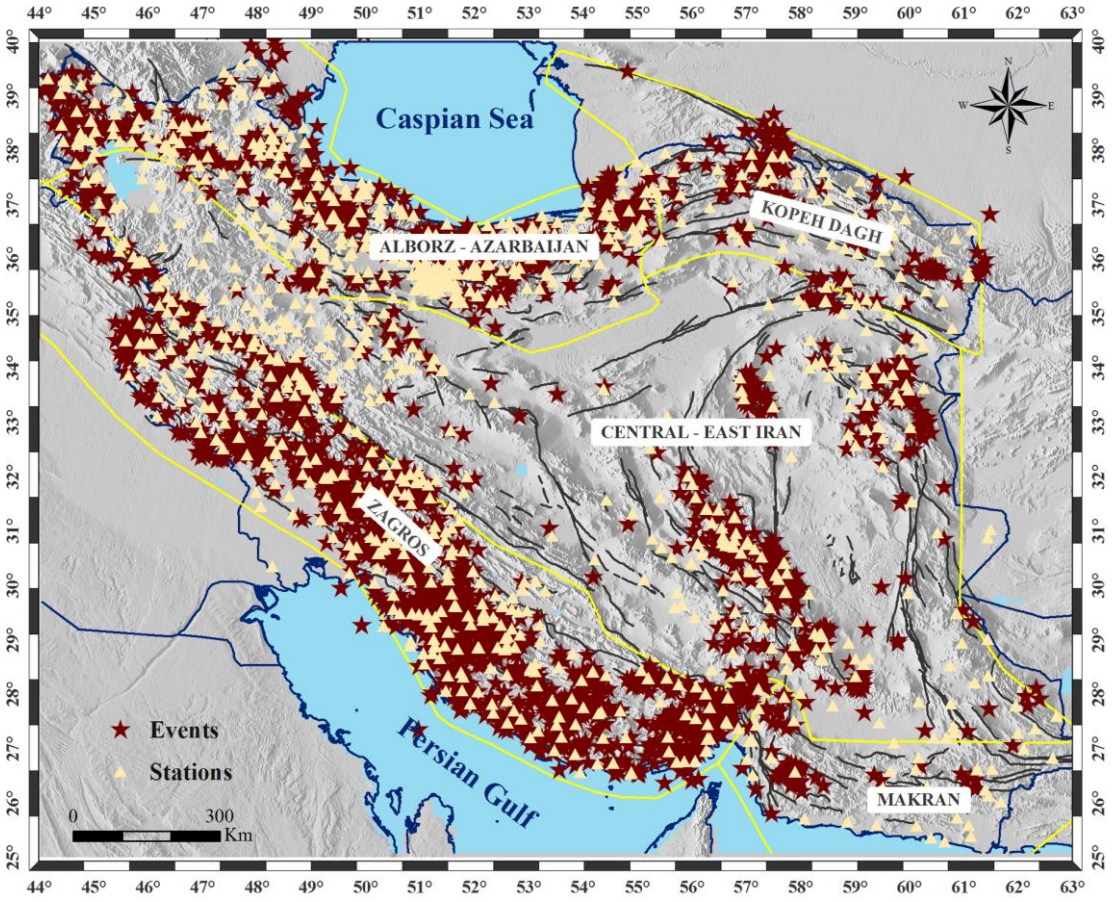


Fig. 7 Location of the strong motion stations and epicenters of the earthquakes (3180 events and 1157 stations)

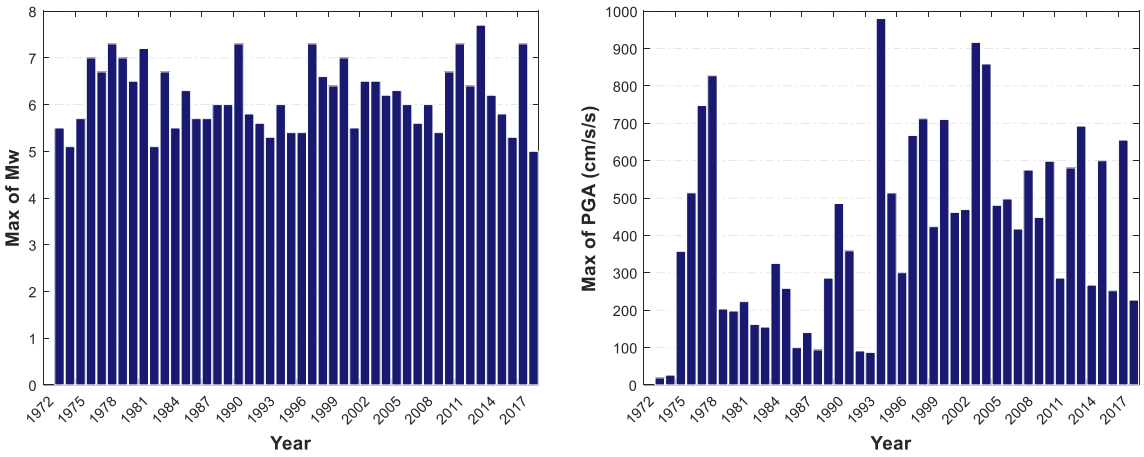


Fig. 8 Distribution of the maximum moment magnitude (*left*) and maximum recorded *PGA* (*right*) in each year since 1973

In this study, the databank is composed of 7196 three-component ground motion records from 3180 earthquakes recorded at 1157 strong motion stations, with moment magnitudes varying from 2.4 to 7.7, occurring in Iran from 1973 to the beginning of 2018. Figure 7 demonstrates the distribution of 1157 strong motion stations, which have recorded 7196 acceleration time series in the selected databank, as well as the epicenters of the selected earthquakes (3180 events). The proposed model of Mirzaei et al. (1998) is used to characterize the seismotectonic and seismicity features of Iran in this work.

The maximum moment magnitude and the maximum *PGA* in each year are shown in Fig. 8. According to this figure, there is at least one earthquake with a moment magnitude greater than 5 in each year in the databank, Furthermore, the figure shows that at least 21 time series accelerations were recorded with *PGA* > 300 cm/s^2 since 1994 (about 25 years).

The histogram of the moment magnitude of the earthquakes and the histogram of the *PGA* of the ground motion records in the databank according to five distinct tectonic regions are shown in Fig. 9. Among 54 earthquakes with $M_w \geq 6.0$ in the databank, the most occurred in Central–East Iran (19 events). Zagros, Alborz–Azarbaijan, Makran, and Kopeh Dagh regions lie in the next ranks with contributions of 17, 12, 4, and 2 events, respectively. Among 91 records with *PGA* > 300 cm/s/s, as exceptional records in the Iranian databank, the most occurred in Zagros region (45 records). The Alborz–Azarbaijan, Central–East Iran, Kopeh Dagh, and Makran regions lie in the next ranks.

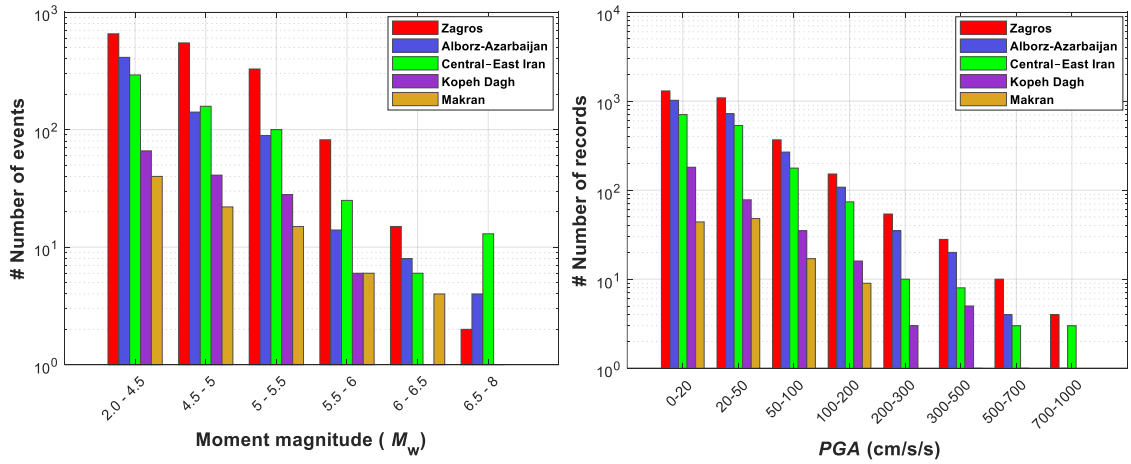


Fig. 9 The histogram of the M_w of earthquakes (*left*) and the histogram of the *PGA* of ground motion records (*right*) in the databank according to five distinct tectonic regions

The geographic distributions of earthquakes in terms of moment magnitude, focal depth and style-of-faulting are shown in Fig. 10. In the selected dataset, as shown in Fig. 10a, the greatest earthquakes occurred in Central Iran (five earthquakes with $M_w \geq 7$). The earthquakes with focal depth exceeding 30 km occurred mostly in the Zagros region (Fig. 10b). Figure 10c also shows that most of the events whose focal mechanisms have been determined in the databank correspond to earthquakes with reverse faulting.

Figure 11 shows the histograms of moment magnitude, epicentral distance, focal depth, and style-of-faulting. The majority of the data are in the magnitude range 2.0-4.5 (about 46% of events), highlighting the dominance of small-size earthquakes (Fig. 11a). A negligible amount of data (about 2% of events) is also available with $M_w \geq 6.0$.

Most of the data are relative to distances between 10 to 30 km (about 40% of the records, 2837 of 7196); 14% and 77% (about 1012 and 5536 of 7196 records) of the accelerogram data were recorded at epicentral distances less than 10 km and 60 km, respectively (Fig. 11b). About 10% of the ground motion records are epicentral distances larger than 120 km (about 694 of 7196 records).

The distribution of recordings in terms of depth intervals (Fig. 11c) shows that most of the earthquakes have focal depth less equal than 30 km, corresponding to about 89% of the total events (about 2831 of 3180 events) and 89% of the ground motion records (6441 of 7196 records), thus indicating a predominance of shallow crustal events in the dataset. Indeed, about 755 records (11%) correspond to deeper events. The events with focal depth more than 30 km are mainly from the Zagros region (Fig. 10b).

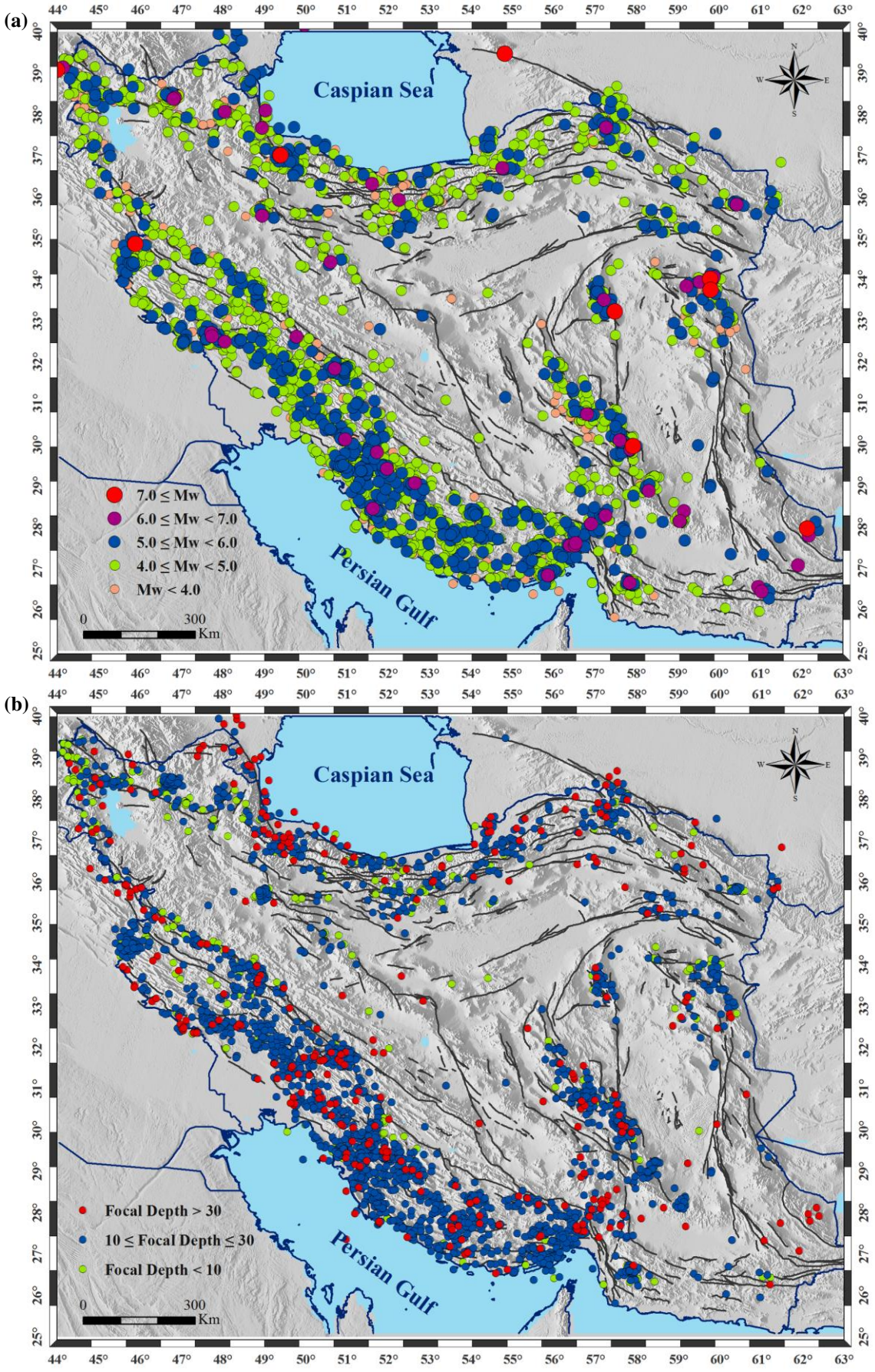


Fig. 10 Geographic distribution of the earthquakes based on **a** moment magnitude, **b** focal depth, and **c** style of faulting

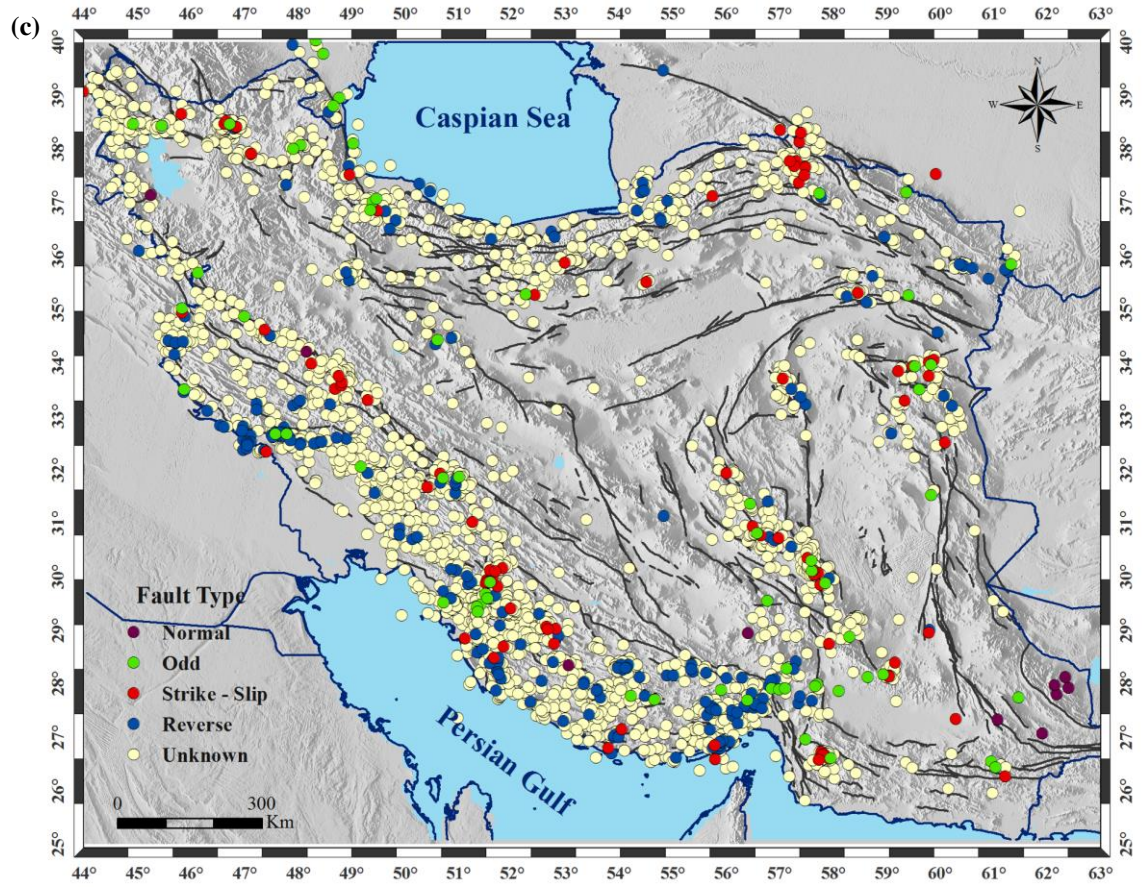


Fig. 10 continued

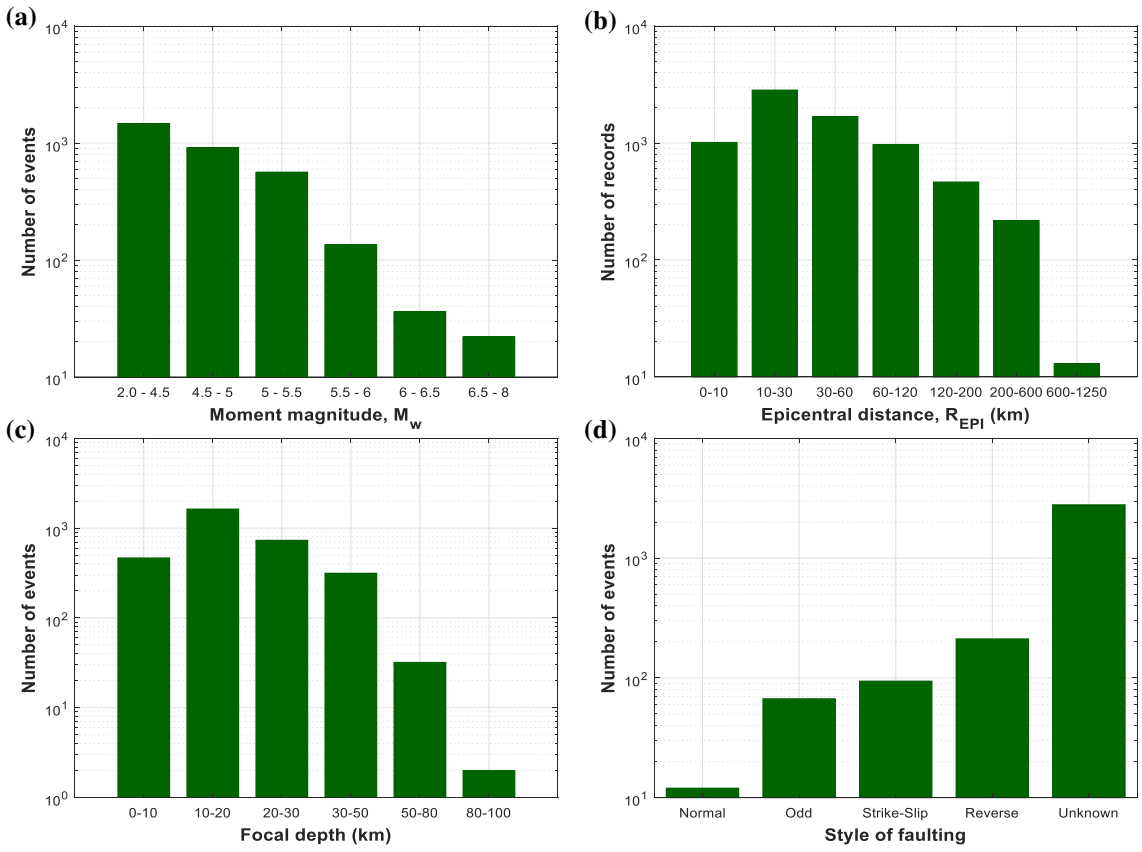


Fig. 11 Histogram of the **a** moment magnitude, **b** epicentral distance, **c** focal depth and **d** style of faulting

Looking at the focal mechanisms distribution in Fig. 11d, the focal mechanisms have been provided for the about 12% of the total number of earthquakes (385 of 3180 events); the earthquakes associated to reverse faulting are predominant (55% of 385 events) in comparison to strike-slip (24% of 385 events), odd (17% of 385 events) and normal (4% of 385 events). In the selected database for most of the earthquakes, about 88% of the events (2795 of 3180 events) and 65% of the ground motion records (4670 of 7196 records), there is no available information about their faulting mechanisms.

The number of earthquakes and ground motion records in terms of faulting mechanisms for each tectonic region is shown in Table 1, respectively. Most of the unknown faulting mechanisms (90% of 2795 events) are attributed to earthquakes with small magnitude ($M_w \leq 5$) due to lack of double-couple fault-plane solutions for these earthquakes. Great events were studied to determine the faulting mechanisms due to their importance, while the small events are rarely paid attention.

Table 1 Number of earthquakes and ground motion records in terms of faulting mechanisms in the five tectonic regions

Region	Number of Earthquakes (Ground motion records)					
	NM	***Od***	***SS***	***RV***	***NA***	***Sum***
Zagros	1 (3)	21 (82)	30 (139)	151 (665)	1446 (2127)	1649 (3016)
Alborz–Azarbaijan	1 (2)	17 (136)	16 (170)	25 (448)	627 (1436)	686 (2192)
Central–East Iran	9 (84)	19 (122)	28 (230)	22 (249)	528 (842)	606 (1527)
Kopeh Dagh	0 (0)	4 (25)	11 (47)	11 (76)	126 (194)	152 (342)
Makran	1 (4)	6 (11)	9 (25)	3 (8)	68 (71)	87 (119)

NM normal, *Od* odd, *SS* strike-slip, *RV* reverse, *NA* unknown.

5 Strong motion stations

The Iranian Strong Motion Network (ISMN) of the Road, Housing and Urban Development Research Center (BHRC) started its operation in 1973 and was equipped first with Kinematics SMA-1 analog strong motion accelerograph in the early years and then have been progressively replaced by SSA-2 digital instruments after the Rudbar earthquake (1990). The SMA-1 instruments recorded important earthquakes such as Tabas (1978) and Golbaf (1981) events. Moreover, the digital strong motion accelerometer (CMG-5TD), solid-state recorder (SSR1), and portable drum recorder (PS2) have been recently added to ISMN.

In this research 1157 stations provide the accelerometric data in the databank. The information for each station involves name, instrument type, coordinates, altitude, azimuth, instrument period and damping, list of identifiers of ground motion records recorded by the station, mean response spectral ratio (H/V) for each station, and site conditions. The first six parameters were obtained from the header information of records provided by the BHRC site (https://smd.bhrc.ac.ir/Portal/en/Search/Waveforms).

In the developed station databank, there are 166, 946, and 45 analog SMA-1, digital SSA-2, and Guralp instruments, respectively. Additionally, 651, 6375, and 170 accelerograms out of a total of 7196 records were recorded by SMA-1, SSA-2, and Guralp instruments, respectively. It is worth noting that the Global Positioning System (GPS) equipment can be installed on the Guralp digital instruments for timing purposes, consequently, the accelerograms recorded by these devices have absolute time. Unfortunately, according to the statistics presented above, 97% of records in Iran's databank do not have absolute time.

The majority of the strong motion stations in Iran, similar to other databases, have limited geotechnical information. Commonly, site conditions are determined using the V_{s30} as a variable in the evaluation of empirical relationships to predict nonlinear (i.e., amplitude-dependent) amplification factors for 5% damped response spectral acceleration. The V_{s30} in more than 50% of the stations in the databank, 603 out of 1157 stations, was measured and reported by the BHRC site (https://smd.bhrc.ac.ir/Portal/en/Search/Stations).

The distribution of the number of strong motion stations versus the number of ground motion records is shown in Fig. 12. The number of records in each station is important in some cases such as single-station sigma studies. In the figure, all stations are distinguished from those with known site classes based on the measured V_{s30}. As shown in this figure, many stations (310 and 273 of 1157 stations) in the databank recorded three to five and one ground motion records, respectively. As well, among the stations with measured V_{s30}, many stations (180, 126, and 125 of 603 stations) recorded three to five, six to ten, and more than ten records, respectively.

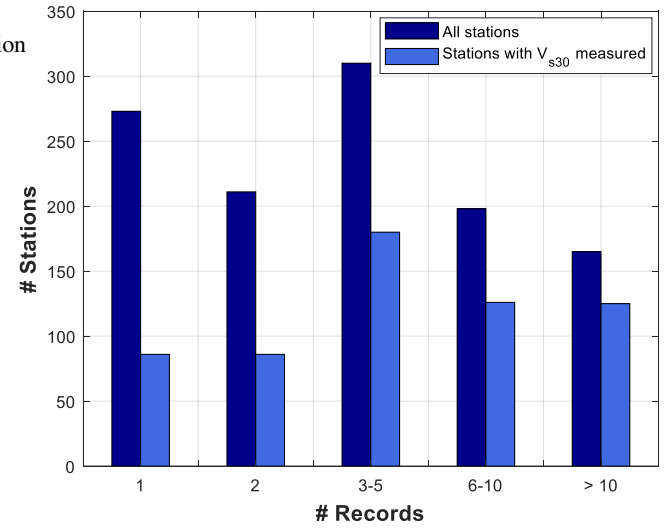


Fig. 12 Histogram of the number of strong motion stations versus number of ground motion records

Many strong motion networks lack measurement of V_{s30}, but it is common to estimate the V_{s30} range from the site characteristics to categorize sites into different general classes. Many researchers use particular solution techniques to estimate V_{s30}. In this study, we used three empirical methods for the estimation of the V_{s30} and site classification stations where there are no measured data. In this regard, first the stations with site classes, previously determined by local geological conditions and V_{s30} measurements, are re-classified and then the correct classification rate of each method is determined and compare with others. The site classification system based on the V_{s30} of the soil column is defined according to the Standard No. 2800: I- $V_{s30} > 750$, II- $375 < V_{s30} < 750$, III- $175 < V_{s30} < 375$ and IV- $V_{s30} < 175$ m/s, which is similar to the one presented in NEHRP (2000).

The method proposed by Wald and Allen (2007) to classify sites using the topographic slope as a proxy to V_{s30} as a primary method is considered. This method develops two sets of coefficients for deriving V_{s30}: one for active tectonic regions with dynamic topographic relief and one for stable continental regions where changes in topography are more smoothed.

As a secondary approach is considered the proposed empirical method of Ghasemi et al. (2009b). Ghasemi et al. a site index, based on the Spearman's rank correlation coefficient, introduced for quantitative site classification using the empirical H/V spectral ratio. Based on this method, the strong motion stations were classified using the average H/V response spectral curves for different site classes obtained in the research of Ghasemi et al. (2009b) and site index proposed by them. This methodology cannot be used for IV site classification.

In the third method, which is actually Ghasemi's updated method, the 603 stations with a total of 4629 recorded strong motions in the databank, with the previously determined V_{s30} parameter are classified into four site categories. The number of stations and records in each group is specified in Table 2.

Table 2 Number of strong motion stations and ground motion records in each site class

Site classes	No. of stations in the databank	No. of registered records	NEHRP class
SC-I	202	1777	B
SC-II	294	2154	C
SC-III	101	684	D
SC-IV	6	14	E

The records of each site category are grouped and used to derive the horizontal to vertical spectral ratio of the motions. In this study, the orientation-independent geometric mean, using the period-independent rotation angle (GMRotI) (Boore et al. 2006), is applied to combine the response spectra of two orthogonal horizontal components. GMRotI is measured for the 50th percentile in each motion. Average H/V response spectral ratios of different site classes are shown in Fig. 13. The amplitudes and shapes of H/V curves are significantly different in different site classes. In this study, a new parameter of the correlation coefficient between the H/V response spectral ratios of the ground motion records recorded by each station has been used for using the appropriate stations in deriving the average H/V response spectral ratios of each site category. By optimization calculations, the appropriate correlation coefficient threshold for site classes I, II, and III was determined as 0.71, 0.607, and 0.65, respectively. No threshold was considered for site class IV. Eventually, stations are classified using the site index proposed by Ghasemi et al. (2009b) and the average H/V response spectral ratios obtained for each site class.

In order to check the applicability of the three proposed methods, the stations with previously determined V_{s30} are re-classified. The overall success rates of the classification schemes for different site classes are shown in Fig. 13. The accuracy rates of these methods are method 1: 29%, 68% and 35%, method 2: 50%, 32% and 57% and method 3: 52%, 34% and 63% for SC-I, SC-II and SC-III, respectively. The overall correct classification rate of method 1 is the highest accuracy for the SC-II site, while this figure shows the higher accuracy of method 3 for SC-I and SC-III sites in comparison with success rates of methods 1 and 2, as well shows higher accuracy of the method 3 for all site classes in comparison of the method 2.

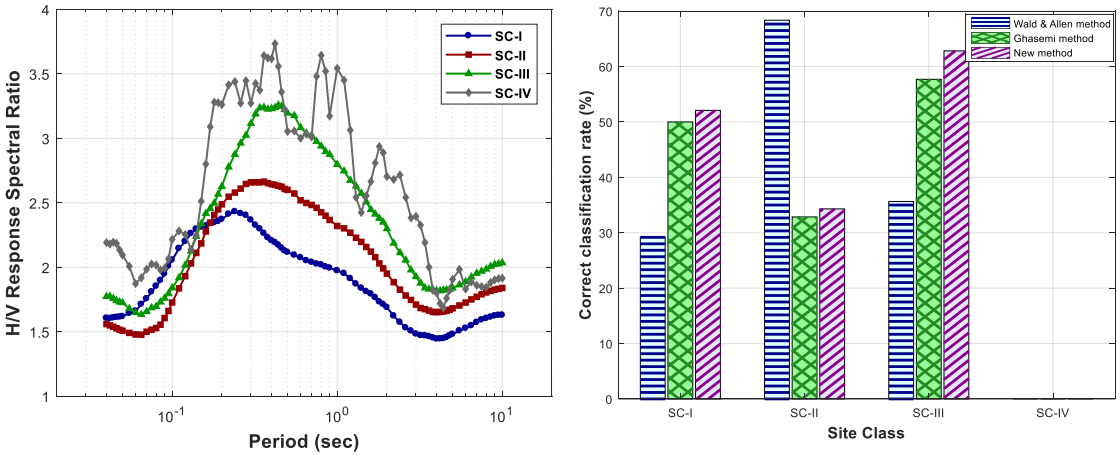


Fig. 13 Average H/V response spectral ratios for each site class (*left*), success rates of the classification schemes (*right*)

For further applicability control, the proposed three schemes are also used to classify a large number of strong motion stations in the databank. An estimation of V_{s30} is provided using method 1 for all stations, while, methods 2 and 3 are used for stations where three or more recorded accelerograms (673 strong motion stations in the databank). Fig. 14d shows the average H/V response spectral ratio for all strong motion station site classes developed in this study, the stations are classified following the third methodology. The mean H/V spectral ratio curves obtained from the research of Ghasemi et al. (2009b), site classification based on the measured V_{s30} (in Fig. 13) and site classification based on the new method proposed in this study are plotted in Fig. 14a, b and c for SC-I, SC-II, and SC-III sites, respectively. The results of each site class in terms of the amplitude and peak period are fairly consistent with those obtained by Ghasemi et al. (2009b) and the ones shown in Fig. 13.

Figure 15 illustrates the histograms of ground motion records and strong motion stations in terms of site class based on measured V_{s30}, estimated V_{s30} by the method of Wald and Allen (2007), estimated site class by the method of Ghasemi et al. (2009b), and estimated site class by the new method in this study. According to this figure, most of the records and stations are in site classes II, I, and III based on the Iranian standard code, using the V_{s30} measured and Wald and Allen empirical method (method 1), respectively. While by estimation of site class using the Ghasemi method and new method, most of the records and stations are in site classes III, II, and I, respectively (methods 2 and 3). Also, it can be seen that in the Iranian strong motion databank, the number of records and stations in site class IV is very rare. Moreover, for 48% of stations and 36% of records, there is no measured V_{s30}, and for 42% of stations and 10% of records, there is no information about soil categories based on methods 2 and 3.

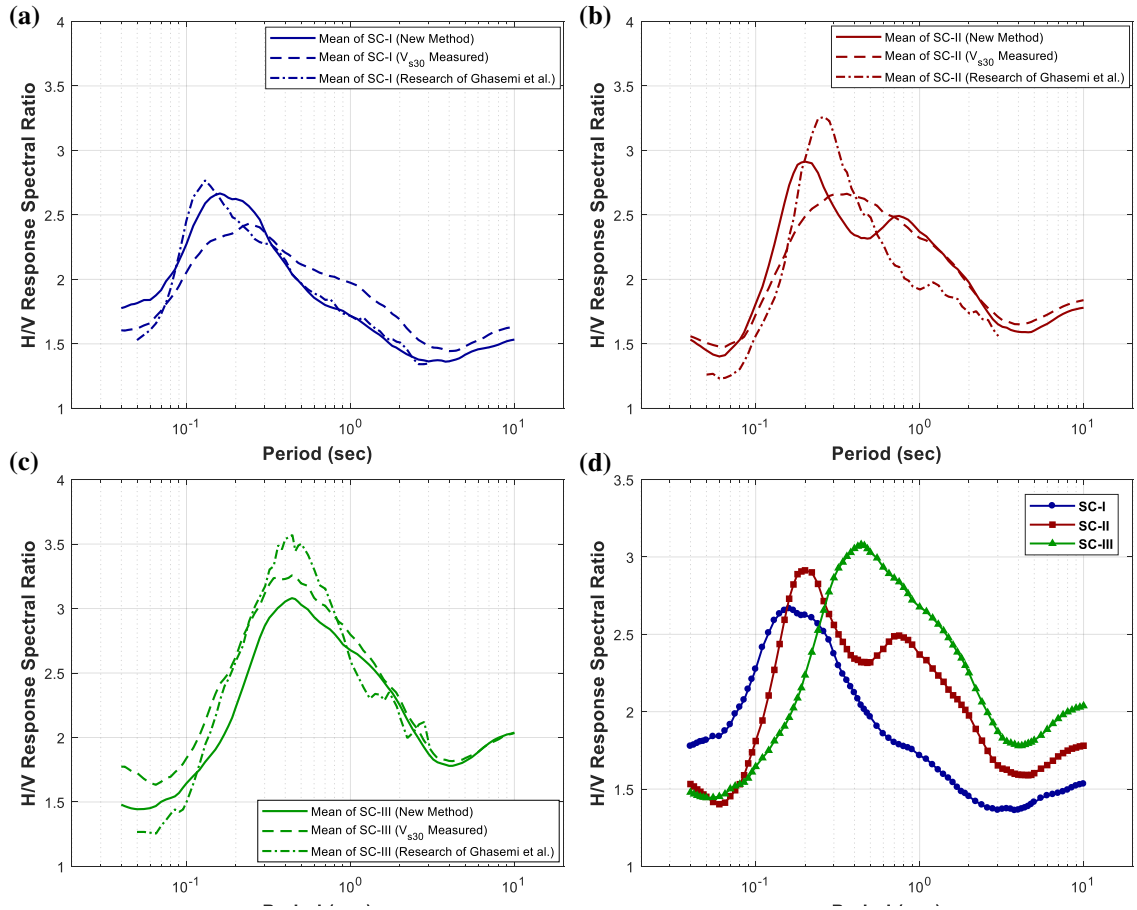


Fig. 14 a-c Comparison between the mean H/V response spectral ratio curves obtained from this study, based on the measured V_{s30} and determined by Ghasemi et al., **d** average H/V response spectral ratios for the site classes assigned based on the new method in this study

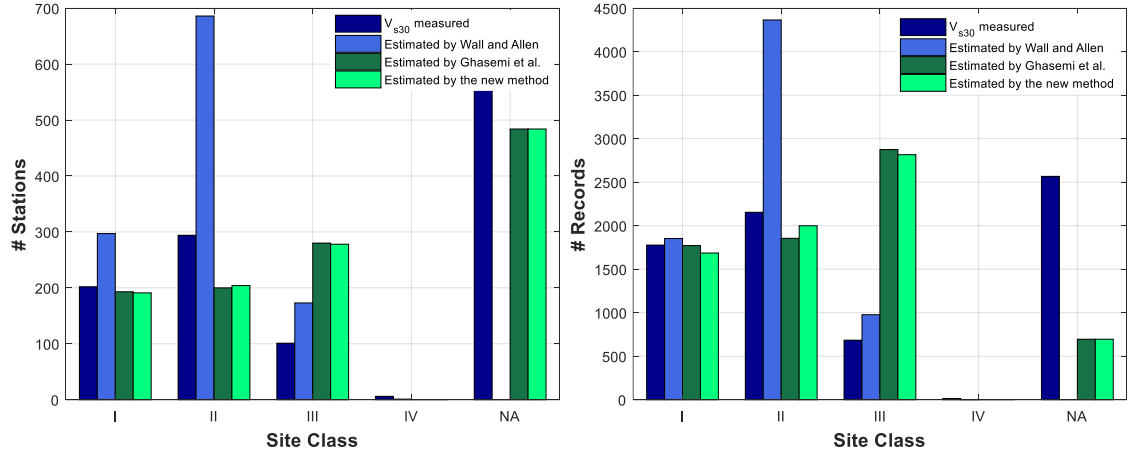


Fig. 15 Distribution of the number of strong motion stations (*left*) and number of recorded ground motion records (*right*) for different site classes based on the Iranian standard code (Standard No. 2800)

Figure 16 shows the spatial distribution of the V_{s30} and the number of accelerograms recorded at stations in the provided databank. In Fig. 16, at the stations where the measured V_{s30} is available, it is used, otherwise, the V_{s30} estimated by Wald and Allen method (Wald and Allen 2007) is used. As shown in this figure, most of the stations located in the Zagros region have higher V_{s30} than the stations located in the Alborz–Azarbaijan region, and most of the stations have V_{s30} between 200 – 1000 m/sec. In addition, the stations in the Makran region have recorded less number of accelerograms.

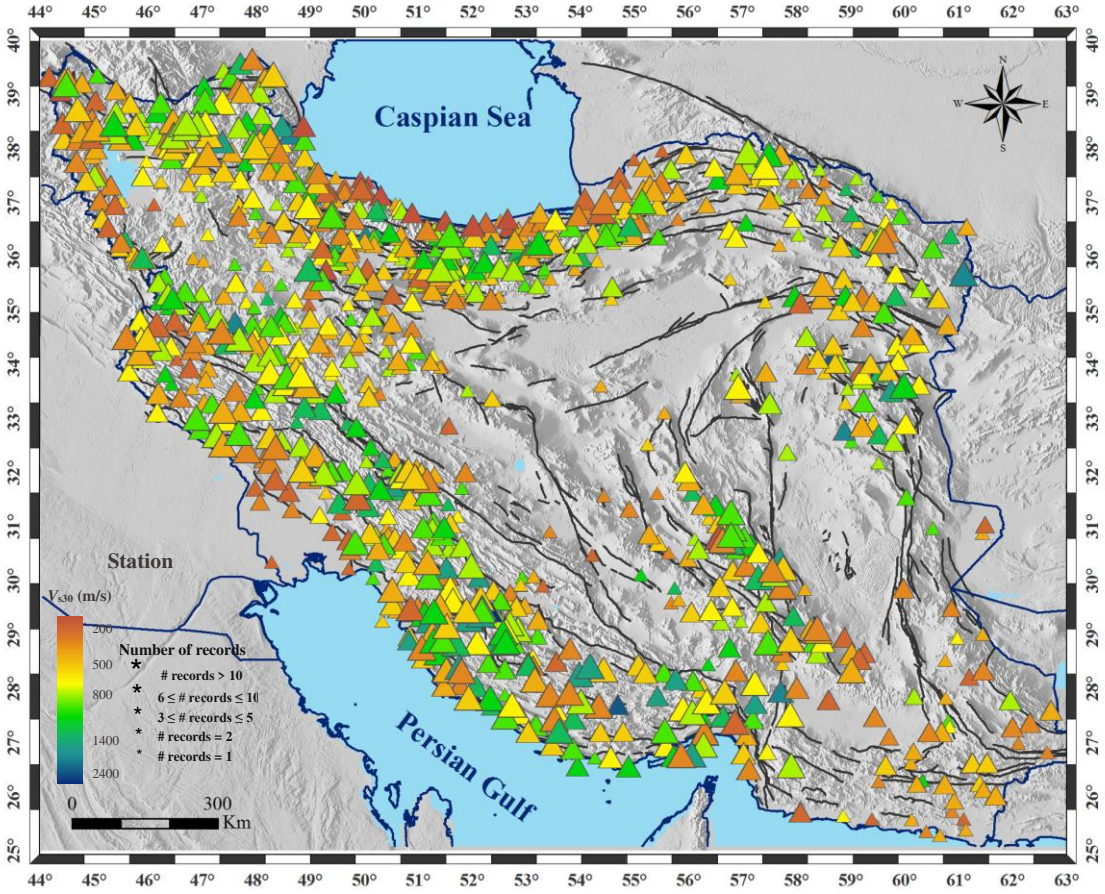


Fig. 16 Spatial distribution of the V_{s30} and the number of accelerograms recorded at recording stations in the developed databank

6 Ground motion records

The ground motion record information for each entry of the databank involves the processed (wavelet de-noising and conventional filtering methods) and unprocessed acceleration time series, filtering corner frequencies, and the various source-to-site distance metrics, namely the epicentral distance (R_{EPI}), hypocentral distance (R_{HYP}), the closest horizontal distance to the vertical projection of the fault rupture plane, named Joyner-Boore distance (R_{JB}) and the closest distance to the fault rupture plane (R_{RUP}). The provided catalog consists of 7196 three-component accelerograms, from 1973 to the end of 2017. The acceleration time series are provided with sampling intervals of 0.005 s in raw (unprocessed) format by BHRC site (https://smd.bhrc.ac.ir/Portal/en/Search/Waveforms).

All information from ground motion records are valuable; therefore, we have tried to obtain all useful information possible from accelerograms which have not be contaminated by noise. Consequently, the uncorrected accelerograms recorded by a given station must be corrected for the instrument response, baseline shifts and high- and low-frequency noises (Boore et al. 2002; Boore and Bommer 2005). To correct the instrument response, the response of the instruments was convolved with the recorded signals. In the case of digital accelerograms, because the natural frequency of transducer is more than the frequency range of interest and damping ratio is properly selected, there is always no need for instrument correction (Boore and Bommer 2005).

The methodology of the baseline correction and the approach to filter out high- and low-frequency noises were studied by several authors (e.g., Boore et al. 2002; Boore and Bommer 2005; Kamai et al. 2014). In this study, was performed a baseline adjustment including subtracting the mean from the uncorrected waveform and removing the linear trend.

The next step is to filter out the noise (conventional filtering method), which can have a substantial impact on the velocity and displacement time histories obtained by the integration of an acceleration time history. It is

practically impossible to separate the noise from the signal; however, data processing and filtering can be used to select part of the signal, which has satisfactory signal to noise ratio (SNR) (Boore and Bommer 2005). The important parts of the filtering procedure are the selection of filtering technique, filter parameters and the usable range of corner frequencies after filtering. The technique based on the signal to noise ratio is employed to find out the corner frequencies, this technique is more suitable for records with long pre-event portions (Boore and Bommer 2005).

A filter is a device or process that, completely or partially, suppresses unwanted components or features from a signal. This usually means removing some frequencies to suppress interfering signals and reduce background noise. The frequency response can be classified into several different band forms describing which frequency bands the filter passes (the passband) and which it rejects (the stopband), such as low-pass filter, high-pass filter, band-pass filter and band-stop filter. In earthquake engineering literature, users can use a wide range of filters such as Butterworth, Chebychev, Bessel, Ormsby and Elliptical, but the main issue is how to use them and their correct application.

Although the choice of filter type is less important, the way in which the filter is applied to the accelerogram has been shown to be very important. The fundamental choice is between causal and acausal filters, the distinguishing feature of the latter being that they do not produce any phase distortion in the signal, whereas causal filters do result in phase shifts in the record. The zero phase shift is achieved in the time domain by passing the transform of the filter along the record from start to finish and then reversing the order and passing the filter from the end of the record to the beginning. The reason that the filters are described as acausal is that to achieve the zero phase shift they need to start to act before the beginning of the record, which can be accomplished by adding lines of data points of zero amplitude, known as pads, before the start of the record and after the end of the record (Boore and Bommer 2005). When acausal filters are applied, the pads are a tool of convenience but their retention as part of the processed record can be important, because, if the padded sections are not retained in further processing, in general, there will be an incompatibility of results derived from the filtered accelerations compared with the results obtained when the padded sections are retained (Boore 2005). It should be noted that, although the use of acausal filters leads to good results, due to the addition of long zero pads to the signal, the length of the signal becomes very long, and this becomes problematic in nonlinear time history analysis (Hatefi and Ashtiany 2015).

In this study, a fourth-order acausal Butterworth filter has been applied to the zero-padded acceleration time series to remove portions of accelerograms for which SNRs are not within the threshold of acceptance. The filter cutoff frequencies for each accelerogram are very important and can easily change values of *PGA*, *PGV*, and 5%-damped elastic pseudo-absolute acceleration response spectral ordinates (*PSA*s) (Akkar and Bommer 2006, 2007). To determine these cutoff frequencies, FAS of the signal to FAS of the noise plot has been used, the lower and upper cutoff frequencies are those at which the SNR becomes less than three (Boore and Bommer 2005).

As before stated, the dataset also contains the filtering corner frequencies for each component of the recordings. Figure 17 shows the scatter plot of the mean high-pass and low-pass corner frequencies of three-component records as a function of M_w: the median value of low-pass frequency is about 29.87 Hz and is found to be magnitude independent; the high-pass frequencies are instead dependent on magnitude, showing the average value of 0.1 Hz for $M_w \leq 5$ and about 0.12 Hz at $M_w > 5$, respectively.

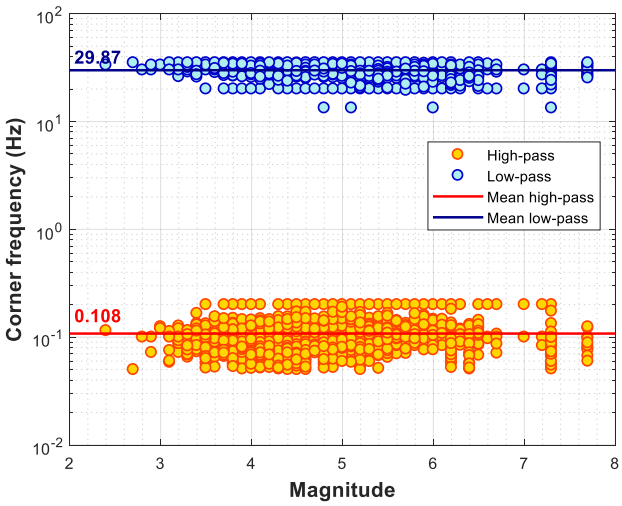


Fig. 17 High-pass and low-pass corner frequencies as a function of M_w

The distribution of recordings in terms of the mean filtering corner frequencies of three components associated with each recording is reported in the databank is shown in Fig. 18. According to this figure, high-pass corner frequency for 61% of the records (4372 of 7196) and low-pass corner frequency for 69% of the records (4974 of 7196) are between 0.1 – 0.12 Hz and 29 - 32 Hz, respectively.

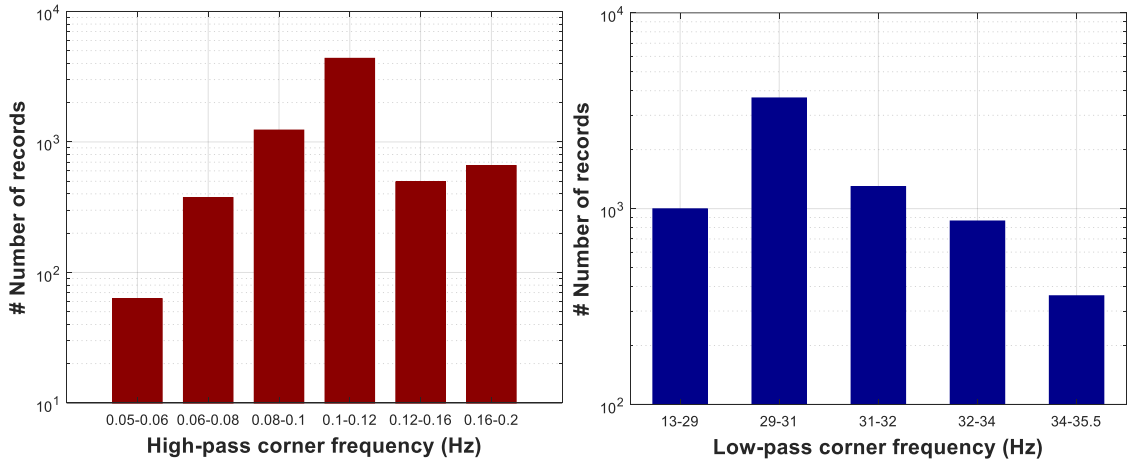


Fig. 18 Histogram of the mean high-pass (*left*) and low-pass (*right*) corner frequencies of three components associated with each recording

In this study, for the processing of the raw time series, in addition to the conventional filtering method, the multi-resolution wavelet de-noising method (Ansari et al. 2007, 2010) has also been used to remove undesirable noise from the recorded signals. Ansari et al. (2007, 2010) proposed a two-step procedure which implements the wavelet de-noising method for the correction of acceleration signals. In the first stage, the correction is applied to acceleration time history and is effective for modification of high-frequency components of the motion. In the second stage, the correction will be applied to the velocity record obtained from the first stage. In this stage, since the moderate and low-frequency components of the motion are amplified due to integration, thus, wavelet de-noising method is more effective in removing the noise of these components. It is shown that in this method, the noise can be attenuated in the whole frequency range of engineering interest. In addition, this method can detect and remove non-stationary noise in the time domain. Ansari et al. (2010) also showed that ‘‘the displacement response spectra of the wavelet de-noised records are more stable than conventional filtered ones and a large number of noisy records that are usually discarded from sets of records used for estimating the ground motions can be corrected using this new method.’’

The histogram of the maximum horizontal peak ground acceleration, velocity and displacement (*PGA*, *PGV* and *PGD*) of the records in the databank according to conventional filtering and wavelet de-noising correction methods is shown in Fig. 19. According to this figure, (a) in both methods most of the records have *PGA* in the range of 10 – 20 (cm/s/s), the number of records with *PGA* less than 10 (cm/s/s) in the wavelet de-noising method is more than the filtering method, while in other intervals the number of records with the wavelet de-noising are almost equal or less than the filtering method; (b) in both methods most of the records have *PGV* in the range of 0.25 – 0.5 (cm/s), the number of records with *PGV* between 0 - 0.5 (cm/s) in the wavelet de-noising method is more than the filtering method, while in other intervals the number of records with the wavelet de-noising are less than the filtering method; (c) in the wavelet de-noising and filtering method most of the records have *PGD* in the range of 0 – 0.05 and 0.05 – 0.1 (cm), respectively.

Fig. 20 shows an example of a horizontal (L) component of a strong motion record from the dataset (7377-1.V1), November 12, 2017, Sarpol-e Zahab earthquake with M_w = 7.3 and R_{EPI} = 68 km, in raw and processed formats. In this figure a comparison between wavelet de-noising and conventional filtering methods is done.

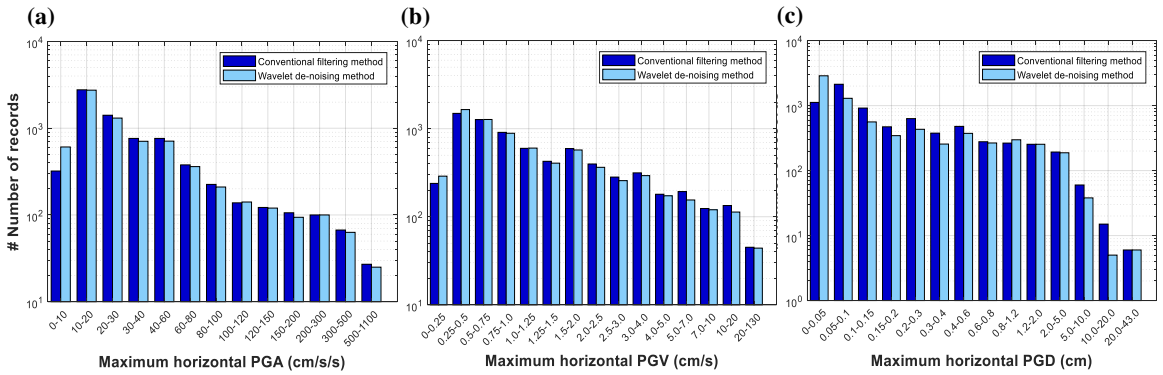


Fig. 19 Histogram of the maximum horizontal: **a** *PGA*, **b** *PGV* and **c** *PGD* of ground motion records in the databank according to conventional filtering and wavelet de-noising methods

In the conventional filtering method, a fourth-order acausal Butterworth filter is used to reduce the noise in the accelerogram. The obtained *PGA*, *PGV* and *PGD* are 308.42 (cm/s/s), 29.84 (cm/s) and 11.78 (cm) in the conventional method, respectively. These parameters are about 308.45 (cm/s/s), 24.99 (cm/s) and 7.13 (cm), respectively, when the wavelet de-noising method is used to process the record (applying eighth-order symlets 8 on the acceleration signal and ninth-order symlets 8 on the velocity signal). The linear elastic displacement, velocity and acceleration response spectra and also the displacement, pseudo-velocity and pseudo-acceleration for the constant-ductility inelastic response spectra for target ductility equal to 2 are shown in the figure for comparison with each other. It is obvious that there is almost no difference between the acceleration response spectra (elastic and inelastic) of the raw record and corrected by the filtering method and corrected by the wavelet de-noising method. The displacement and velocity response spectra (elastic and inelastic) of the record corrected by the wavelet de-noising method have lower values than the raw record and corrected by the filtering method. The linear elastic displacement response spectra of the record corrected by the wavelet de-noising method reaches the constant value of the *PGD* in less time than the raw record and the record corrected by the filtering method.

One of the best analyses that can be done on the developed accelerograms databank is the comparison between FAS of the noises detected by the filtering and wavelet de-noising correction methods from the raw ground motion records of the databank. For this purpose, the acceleration and velocity signals corrected by the filtering and wavelet de-noising methods are subtracted from the raw acceleration and velocity signals, respectively, then the FAS of the detected acceleration and velocity noises is calculated. To better compare the FAS of the noises detected from raw ground motion records, two approaches have been used: in the first method, the average FAS of the noises of the three components of each accelerogram is calculated, and then the mean FAS of the noises in all accelerograms at each frequency is obtained. In the second method, after calculating the average FAS of the noises of the three components of each accelerogram, the value that appears to have the most frequent in all accelerograms at each frequency is calculated (mode of the data set). The results of the two proposed approaches for the acceleration and velocity signal noises are shown in Fig. 21. As can be seen from this figure, the difference between the FAS of the detected noises (by two correction methods) in the acceleration signal is more than the velocity signal, also, this difference in the case where the most frequent value in the dataset at each frequency is used is more than in the case where is used the mean of the dataset at each frequency.

According to Fig. 21a and 21b, the mean FAS of the acceleration noises detected by the wavelet de-noising method in the frequency range of 0.1 to 25 Hz is at least 1.2 times and at most 5 times the corresponding value obtained from the filtering correction method and the mean FAS of the acceleration noises detected by the filtering method in the frequency range of 25 to 80 Hz is at most 1.7 times the corresponding value obtained from the wavelet de-noising correction method, in other frequencies, the two graphs are identical. Whereas, the mean FAS of the velocity noises detected by the two methods was almost identical in all frequencies.

Figures 21c and 21d show the difference between the FAS of the acceleration and velocity noises detected by the two correction methods much more than figures 21a and 21b, so that the wavelet de-noising correction method shows the FAS of the acceleration and velocity noises in some frequencies up to 23 times and 22 times of the filtering correction method, respectively. As well as the most difference between the two graphs (two correction methods) occurred at the frequency of 3.36 Hz.

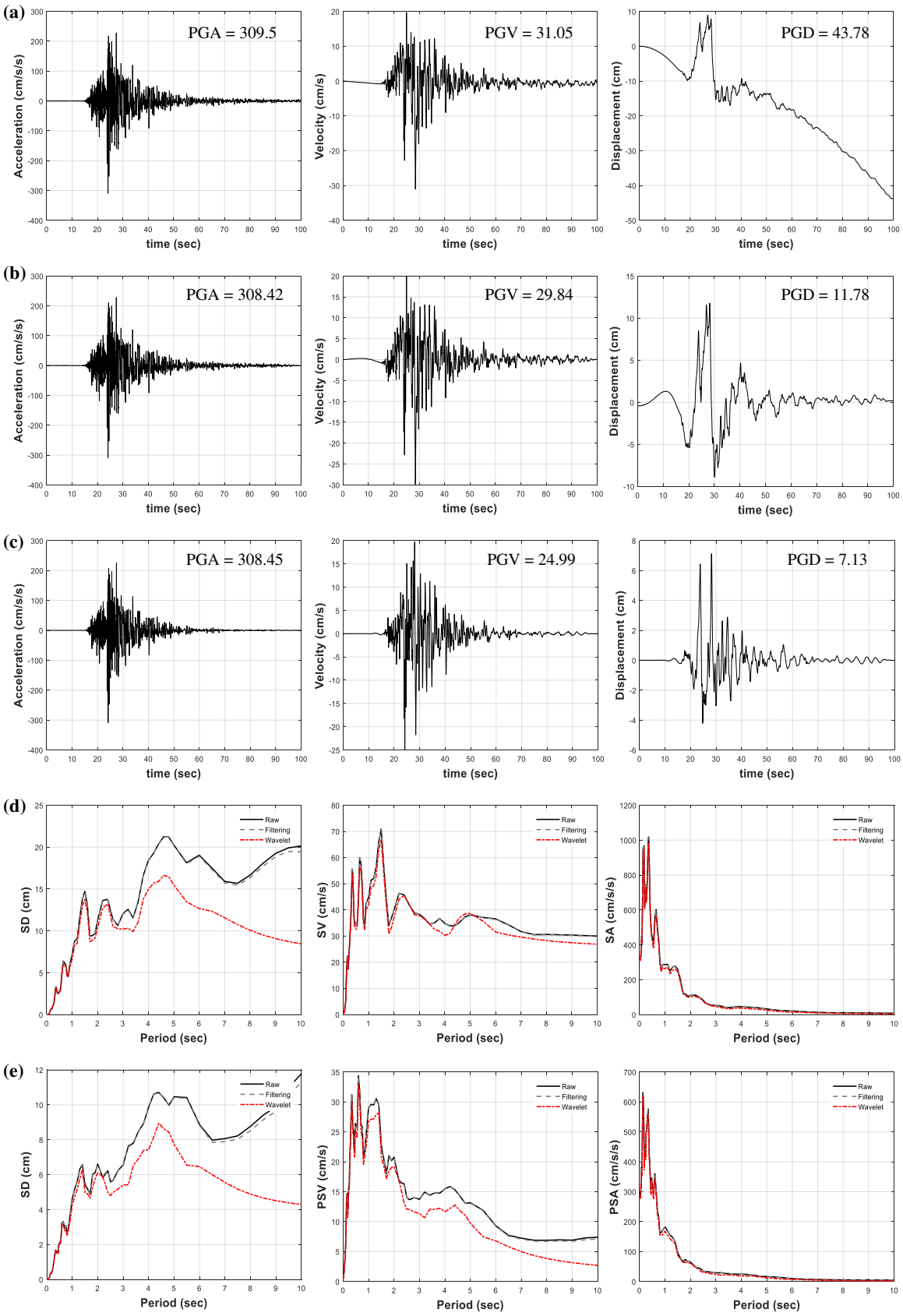


Fig. 20 a One sample ground motion record in raw format, **b** processed by the conventional filtering method, **c** corrected by the wavelet de-noising approach, **d** comparison of the linear elastic spectral amplitudes and **e** comparison of the constant ductility inelastic spectral amplitudes

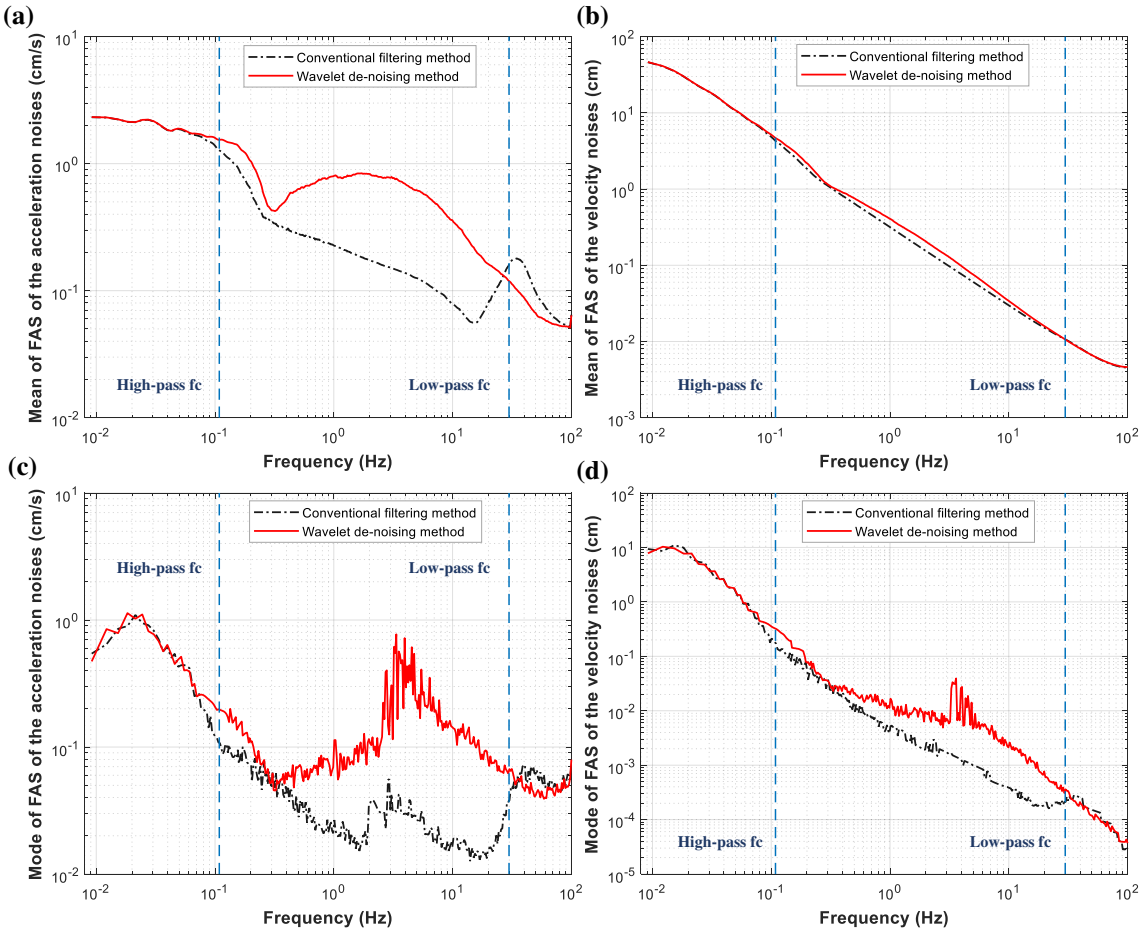


Fig. 21 **a** Mean FAS of the acceleration noises, **b** Mean FAS of the velocity noises, **c** The most frequent value of FAS of the acceleration noises and, **d** The most frequent value of FAS of the velocity noises at each frequency

It is necessary to mention that, in the filtering method, some low and high frequency components of the motion are removed from the signal, while other frequency components of the signal remain unchanged because it is assumed that the energy of the noise is concentrated only in the low and high frequencies; however, in applying the wavelet de-noising method, it is assumed that it is possible to have noise in all frequency components of the motion, just like the case of white noise.

All steps of processing and correcting the accelerograms of the databank, as well as calculating all intensity measure parameters have been done using the IIEES toolbox. The IIEES toolbox is a comprehensive tool developed for data processing of ground motion records, producing the design spectra, elastic and inelastic spectrum, calculating the engineering intensity measure parameters and also with capability of corrected record selection (the manuscript is in preparation).

The other important information given for each ground motion record is the source-to-site distances that are measured from various references. The source-to-site distances include R_{EPI} for all the records, R_{HYP}, R_{JB} and R_{RUP}. The readers are referred to Abrahamson and Shedlock (1997) for the generic definitions of these distance metrics. In this study, for the calculation of R_{JB} and R_{RUP}, the main fault-plane for each event is determined based on the double-couple fault-plane solutions extracted from the Harvard Global Centroid Moment Tensor (GCMT) catalog international agency. For such cases, upon the existence of double-couple fault-plane solutions, the rupture dimensions of the fault (length and width) are estimated from Wells and Coppersmith (1994) and for the determination of the relative location of the hypocenter with respect to fault dimensions, we have used the analysis results of the hypocenter point in finite-source rupture models with respect to the overall fault dimension presented by Mai et al. (2005). The framework of Kaklamanos et al. (2011), which is suggested for estimating the unknown source, source-to-site distance, and site parameters when implementing the NGA-West project (Ancheta et al. 2014), is used as a reference for calculating missing parameters in our databank.

One of the sources of error in calculating the source-to-site distance is the incorrect selection of the fault plane from the moment solution. In Figure 22, R_{JB} and R_{RUP} in pairs for the main and alternate planes of the fault are shown. The observations on the computed R_{JB} pairs indicate that differences between the components of each pair are small for far-source accelerograms and earthquakes with M_w greater than 5.5. The difference between the components of R_{JB} pairs becomes significant for some recordings with M_w between 4.5 to 6.5 that are close to the source. Based on the computed R_{RUP} pairs the differences between the components of each pair are small for the majority of accelerograms. As well, the difference between the components of R_{RUP} pairs is significant for some large-magnitude ($6.5 \leq M_w \leq 7.7$) recordings that are far from the source.

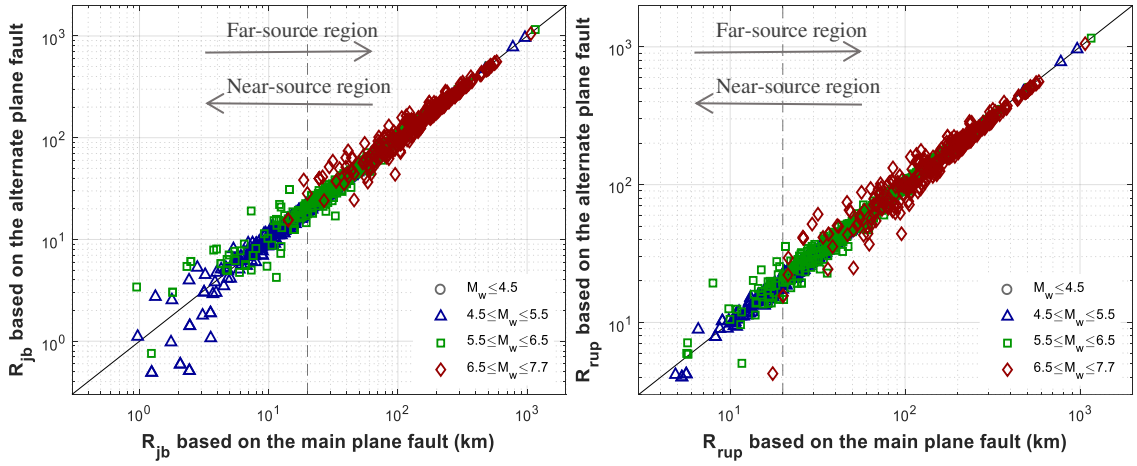


Fig. 22 Differences between R_{JB} (*left*) and R_{RUP} (*right*) pairs computed from the two main and alternate planes of the fault given by the double-couple fault-plane solutions

The distribution of the M_w versus distance is shown in Fig. 23. It is obvious from the figure, that ground motion records are scarce for site class IV in the databank. The distance metric related to fault geometry (R_{JB}) is computed by additional information about the ruptured fault geometry. Since there is not enough information on ruptured fault geometry and double-couple fault-plane solutions of smaller magnitude earthquakes ($M_w < 4$), R_{JB} is not available for the records with unknown ruptured fault geometry, which are 65% of the ground motion records in the developed databank. For small earthquakes, R_{EPI} and R_{JB} are similar because of the small rupture planes of such earthquakes. Akkar et al. (2010) showed that the R_{JB} is almost smaller than R_{EPI} and the difference between R_{JB} and R_{EPI} can be noticeable for both $M_w \geq 6$ and epicentral distances less than 40 km, which is observed in this study according to this figure.

Figure 23 illustrates the scarcity of data for magnitudes greater than 6 at distances less than 10 km (i.e., 11 of 7196 records) in the databank. In addition, 14% of the records (1011 of 7196 records) have epicentral distances less than 10 km in the entire moment magnitudes range, and the records are dominant in source-to-site distance between 10 and 30 km.

We illustrate the relationships between distance metrics (i.e., R_{JB} vs. R_{EPI}, R_{RUP} vs. R_{EPI}, R_{RUP} vs. R_{JB}, and R_{RUP} vs. R_{HYP}) in Fig. 24 for different moment magnitude intervals. Figure 24a shows that R_{JB} attains smaller than values with respect to R_{EPI}, especially for events with $M_w \geq 5$. The discrepancies between R_{JB} and R_{EPI} diminish at large distances except for a few earthquakes with $M_w \geq 7$. When M_w is greater than 5 and epicentral distances are less than 40 km, the discrepancy between R_{JB} and R_{EPI} becomes noticeable.

The scatters in Fig. 24b reveal that R_{RUP} generally tends to be larger than R_{EPI} for decreasing magnitude and distance, whenever R_{EPI} tends to be larger than R_{RUP} for increasing magnitude and decreasing distance. In the case of R_{RUP} vs. R_{JB} relationship (Fig. 24c), R_{RUP} tends to be larger than R_{JB} for $R_{JB} < 30$ km regardless of the variations in magnitude. For larger distances, R_{JB} is approximately equal to R_{RUP}, which means that above approximately 30-40 km, the differences in the definitions of these distance metrics become immaterial. This can be attributed to the importance of the depth of an earthquake that marks the major differences between R_{JB} and R_{RUP} for sites close to the fault rupture plane. As the recording station is located away from the source, depth loses its significance and consequently $R_{JB} \approx R_{RUP}$. Note that the databank consists mainly of shallow earthquakes, and this feature reinforces our observations on the similarity of R_{RUP} and R_{JB} at the intermediate and large distances.

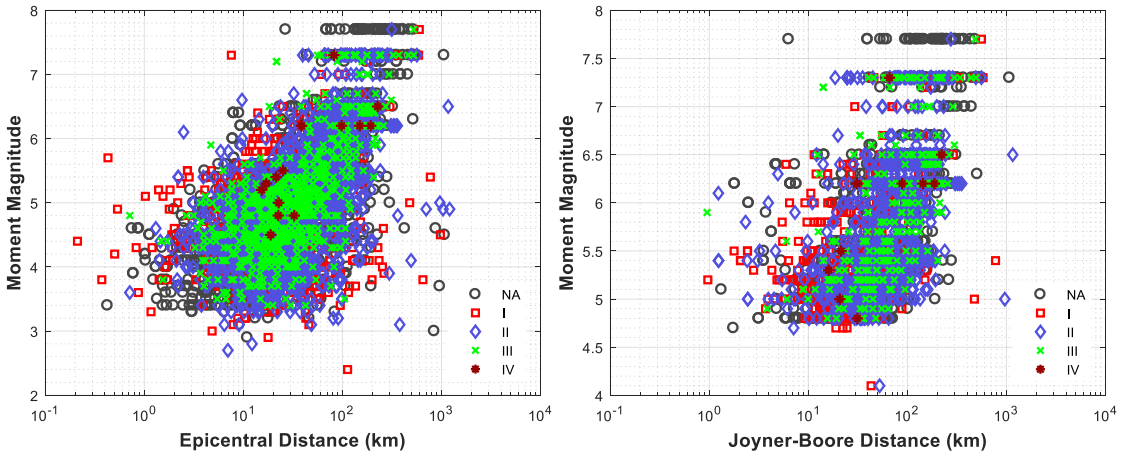


Fig. 23 Distribution of Iranian ground motion databank with respect to moment magnitude, R_{EPI} (*left*) and R_{JB} (*right*) grouped by site class

It is depicted in Fig. 24d that R_{HYP} is almost greater than R_{RUP}, only for $M_w < 5$ and $R_{HYP} < 10$ km, R_{RUP} is greater than R_{HYP}. When the ground motions represent close-to-intermediate distance recordings (i.e., $10 < R_{HYP} < 60$ km), regardless of the variations in magnitude, the discrepancy between these two distance metrics becomes larger.

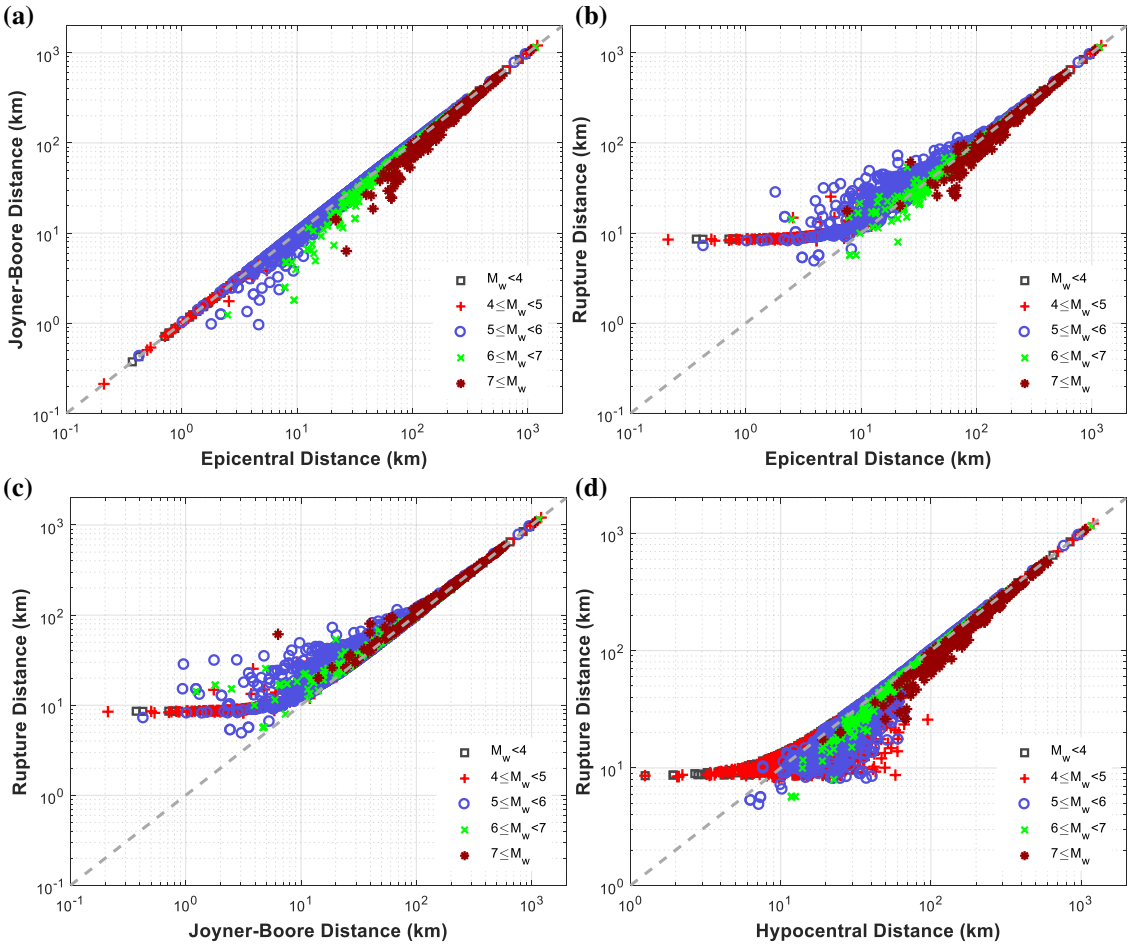


Fig. 24 Magnitude-dependent relationships between various source-to-site distance metrics obtained from the Iranian ground motion databank

Distribution of the engineering intensity measure parameters such as peak ground motion values, response spectra ordinates in acceleration, velocity and displacement, Fourier and effective amplitude spectrum ordinates, numerous definitions of the strong motion duration and characteristic periods and the most common integral parameters (Arias intensity, Housner intensity, cumulative absolute velocity, specific energy density, and …) is discussed in another study that is being prepared.

7 Summary and conclusions

The main purpose of this study is to develop a comprehensive, reliable and unified databank of the Iranian corrected strong ground motions. Generally, the collected databank was developed, including the manually reviewed parameters of earthquakes, stations and ground motion records from international and national authoritative sources (e.g. ISC, GCMT, IRSC and BHRC) and the results of the earthquake-specific literature studies. The current version of the Iranian databank has been increased by approximately 3 times in terms of the size of earthquakes, stations and recorded accelerograms compared to previous similar researches, with improvements and increased accuracy in the distribution of distance and magnitude, site classification and processed and corrected ground motion records.

The IIEES toolbox (the manuscript is in preparation) connects to this developed Iranian ground motion databank to perform a preliminary selection of appropriate ground motion records for earthquake engineering and engineering seismology research based on a set of criteria such as distance, magnitude, V_{s30}, peak ground parameters, and lowest usable frequency.

The current size of Iranian databank consists of 7196 three-component accelerograms from 3180 earthquakes of different tectonics regimes, although mainly related to shallow active crustal regions, recorded by the 1157 strong motion stations between 1973 and 2018. The moment magnitude is provided for almost 12% of events from international and national bulletins. Therefore, other magnitude scales were converted to the M_w scale using the best available correlations to obtain a unified catalog. The earthquake M_w range between 2.4 and 7.7 in the databank. The entire databank has R_{EPI} source-to-site distance information. The corresponding numbers for R_{HYP}, R_{JB} and R_{RUP} source-to-site distance metrics are 7170, 2526 and 2526, respectively. Since one of the sources of error in calculating the source-to-site distance is the incorrect selection of the fault plane from the moment solution, calculated R_{JB} and R_{RUP} distances for the main and alternate planes of the fault were evaluated. The results showed that the differences between the components of R_{JB} pairs are small for far-source accelerograms and earthquakes with $M_w > 5.5$ and become significant for some recordings with $4.5 < M_w < 6.5$ that are close to the source. Based on the computed R_{RUP} pairs, differences between the components of each pair are insignificant for the majority of accelerograms.

All the ground motion records in the databank, which were recorded in raw form by ISMN of the BHRC, were uniformly corrected for the instrument response, baseline shifts and processed using two methods of the conventional filtering and wavelet de-noising procedure. The results of processing raw accelerograms with two correction methods showed that there is almost no difference between the peak parameters of acceleration and acceleration response spectra of the corrected record by the filtering and the wavelet de-noising method. On the other, in most of the records, correction with the de-noising wavelet method causes a greater reduction in the peak parameters of velocity and displacement and also displacement and velocity response spectra compared to the conventional filtering method. From another point of view, the mean and mode of FAS of the acceleration and velocity noises detected at each frequency by the wavelet de-noising method in the band-pass frequency range are much more than the corresponding value obtained from the filtering correction method, and at other frequencies (high-pass and low-pass frequencies range) the FAS of the noises detected by both methods is identical.

Shear wave velocity for the top 30 m of soil (V_{s30}) is another important parameter that we included in the databank. The V_{s30} of 603 of the total strong motion stations was measured and reported by the BHRC. In this present study, three empirical techniques, including Wald and Allen (2007), Ghasemi et al. (2009b) and a new method, are also used to classify strong motion sites. In the new method, which is actually an updated Ghasemi method, stations in the developed databank with the previously determined V_{s30} parameter are classified into different site classes, based on the V_{s30} ranges proposed by Standard No. 2800. The ground motion records of each site category are grouped and used to derive the mean H/V spectral ratios. In this study, a new parameter of the correlation coefficient between the H/V response spectral ratios of the ground motion records recorded by each

station is proposed. By optimization analysis and determining the correlation coefficient threshold, the stations are classified using the site index introduced by Ghasemi et al. (2009b). In order to check the applicability of the proposed methods, the stations with previously determined V_{s30} are re-classified. The success rate of the new method for SC-I, SC-II, and SC-III is 52%, 34%, and 63%, respectively. The success rate of the proposed method is higher than methods 1 and 2 for SC-I and SC-III site classes. It is important to mention that the second and third methods can only be used for strong motion stations with three or more recorded accelerograms.

One of the goals of this dataset is to create a standard for the dissemination of waveform metadata and parameters for engineering seismology and earthquake engineering applications in Iran, similarly to NGA-West in the United States. A significant effort in the compilation of the database has been spent in the metadata revision to make them traceable including specific fields for the references of each metadata (i.e. event, record and station). The data collected in the dataset can be used for the structural design, nonlinear dynamic analysis, development of ground motion prediction equations, development of rapid losses and fatality estimation models, selection appropriate and ranking of existing GMPEs for hazard assessment, calibration of ShakeMap models, performing statistical goodness-of-fit tests and calibration of new GMPEs for different tectonic regimes in Iran and other seismically active environments in the Middle East and Mediterranean regions.

Acknowledgments The authors express their sincere appreciation to the Road, Housing and Urban Development Research Center (BHRC) for providing the recorded ground motion data. The first author wishes to forward thanks to Dr. Erfan Firuzi for the valuable discussions and helpful comments.

References

Abrahamson NA, Shedlock KM (1997) Overview. Seismol Res Lett 68:9–23. https://doi.org/10.1785/gssrl.68.1.9

Akkar S, Bommer JJ (2006) Influence of long-period filter cut-off on elastic spectral displacements. Earthq Eng Struct Dyn 35:1145–1165

Akkar S, Bommer JJ (2007) Prediction of elastic displacement response spectra in Europe and the Middle East. Earthq Eng Struct Dyn 36:1275–1301

Akkar S, Çağnan Z, Yenier E, et al (2010) The recently compiled Turkish strong motion database: preliminary investigation for seismological parameters. J Seismol 14:457–479. https://doi.org/10.1007/s10950-009-9176-9

Ancheta TD, Darragh RB, Stewart JP, et al (2014) NGA-West2 database. Earthq Spectra 30:989–1005

Ansari A, Noorzad A, Zafarani H, Vahidifard H (2010) Correction of highly noisy strong motion records using a modified wavelet de-noising method. Soil Dyn Earthq Eng 30:1168–1181. https://doi.org/https://doi.org/10.1016/j.soildyn.2010.04.025

Ansari A, Noorzad A, Zare M (2007) Application of wavelet multi-resolution analysis for correction of seismic acceleration records. J Geophys Eng 4:362–377. https://doi.org/10.1088/1742-2132/4/4/002

Atkinson GM, Boore DM (2007) Boore-Atkinson NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. Pacific Earthquake Engineering Research Center

Boore D (2005) On Pads and Filters: Processing Strong-Motion Data. Bull Seismol Soc Am - BULL Seism SOC AMER 95:745–750. https://doi.org/10.1785/0120040160

Boore D, Stephens C, Joyner W (2002) Comments on baseline correction of digital strong-motion data: examples from the 1999 Hector Mine, California, earthquake. Bull Seismol Soc Am 92:1543–1560

Boore DM, Bommer JJ (2005) Processing of strong-motion accelerograms: Needs, options and consequences. Soil Dyn Earthq Eng 25:93–115. https://doi.org/10.1016/j.soildyn.2004.10.007

Boore DM, Watson-Lamprey J, Abrahamson NA (2006) Orientation-Independent Measures of Ground Motion. Bull Seismol Soc Am 96:1502–1511. https://doi.org/10.1785/0120050209

Council BSS (2000) The 2000 NEHRP recommended provisions for new buildings and other structures: Part I (Provisions) and Part II (Commentary). FEMA 368:369

Farajpour Z, Pezeshk S, Zare M (2019) A New Empirical Ground-Motion Model for Iran. Bull Seismol Soc Am 109:732–744. https://doi.org/10.1785/0120180139

Farajpour Z, Zare M, Pezeshk S, et al (2018) Near-source strong motion database catalog for Iran. Arab J Geosci 11:1–16

Firuzi E, Amini Hosseini K, Ansari A, et al (2020) An empirical model for fatality estimation of earthquakes in Iran. Nat Hazards 103:231–250. https://doi.org/10.1007/s11069-020-03985-y

Firuzi E, Amini Hosseini K, Ansari A, Tabasian S (2022) Developing a new fatality model for Iran's earthquakes using fuzzy regression analysis. Int J Disaster Risk Reduct 80:103231. https://doi.org/https://doi.org/10.1016/j.ijdrr.2022.103231

Frohlich C, Apperson KD (1992) Earthquake focal mechanisms, moment tensors, and the consistency of seismic activity near

plate boundaries. Tectonics 11:279–296

Ghasemi H, Zare M, Fukushima Y, Koketsu K (2009a) An empirical spectral ground-motion model for Iran. J Seismol 13:499–515. https://doi.org/10.1007/s10950-008-9143-x

Ghasemi H, Zare M, Fukushima Y, Sinaeian F (2009b) Applying empirical methods in site classification, using response spectral ratio (H/V): A case study on Iranian strong motion network (ISMN). Soil Dyn Earthq Eng 29:121–132. https://doi.org/https://doi.org/10.1016/j.soildyn.2008.01.007

Goulet CA, Kottke A, Boore DM, et al (2018) Effective amplitude spectrum (EAS) as a metric for ground motion modeling using Fourier amplitudes. In: 2018 Seismology of the Americas Meeting

Hamzehloo H, Mahood M (2012) Ground-Motion Attenuation Relationship for East Central Iran. Bull Seismol Soc Am 102:2677–2684. https://doi.org/10.1785/0120110249

Hatefi H, Ashtiany M (2015) The Effect of Strong Ground Motion Signal Processing on the Responses of Nonlinear Dynamic Analysis

Kaklamanos J, Baise L, Boore D (2011) Estimating Unknown Input Parameters when Implementing the NGA Ground-Motion Prediction Equations in Engineering Practice. Earthq Spectra 27:1219–1235. https://doi.org/10.1193/1.3650372

Kamai R, Abrahamson NA, Graves R (2014) Adding fling effects to processed ground-motion time histories. Bull Seismol Soc Am 104:1914–1929. https://doi.org/10.1785/0120130272

Konno K, Ohmachi T (1998) Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. Bull Seismol Soc Am 88:228–241. https://doi.org/10.1785/BSSA0880010228

Mai PM, Spudich P, Boatwright J (2005) Hypocenter locations in finite-source rupture models. Bull Seismol Soc Am 95:. https://doi.org/10.1785/0120040111

Mirzaei N, Mengtan G, Yuntai C (1998) Seismic source regionalization for seismic zoning of Iran: major seismotectonic provinces. J Earthq Predict Res 7:465–495

Mousavi-Bafrouei SH, Mahani AB (2020) A comprehensive earthquake catalogue for the Iranian Plateau (400 B.C. to December 31, 2018). J Seismol 24:709–724. https://doi.org/10.1007/s10950-020-09923-6

Naserieh S, Karkooti E, Dezvareh M, Rahmati M (2019) Analysis of artifacts and systematic errors of the Iranian Seismological Center's earthquake catalog. J Seismol 23:665–682. https://doi.org/10.1007/s10950-019-09828-z

Sedaghati F, Pezeshk S (2017) Partially Nonergodic Empirical Ground-Motion Models for Predicting Horizontal and Vertical PGV, PGA, and 5% Damped Linear Acceleration Response Spectra Using Data from the Iranian Plateau. Bull Seismol Soc Am 107:934–948. https://doi.org/10.1785/0120160205

Storchak DA, Harris J, Brown L, et al (2020) Rebuild of the Bulletin of the International Seismological Centre (ISC)—part 2: 1980–2010. Geosci Lett 7:18. https://doi.org/10.1186/s40562-020-00164-6

Wald DJ, Allen TI (2007) Topographic slope as a proxy for seismic site conditions and amplification. Bull Seismol Soc Am 97:1379–1395

Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. Bull Seismol Soc Am 84:974–1002

Willemann RJ, Storchak DA (2001) Data collection at the international seismological centre. Seismol Res Lett 72:440–453

Zafarani H, Luzi L, Lanzano G, Soghrat MR (2018) Empirical equations for the prediction of PGA and pseudo spectral accelerations using Iranian strong-motion data. J Seismol 22:263–285. https://doi.org/10.1007/s10950-017-9704-y

Zafarani H, Soghrat MR (2017) A selected dataset of the Iranian strong motion records. Nat Hazards 86:. https://doi.org/10.1007/s11069-017-2745-2

Zoback M Lou (1992) First-and second-order patterns of stress in the lithosphere: The World Stress Map Project. J Geophys Res Solid Earth 97:11703–11728

Statements and Declarations

Funding This study is conducted as part of a research project being implemented in the International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran. This project is funded by IIEES.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

Author Contributions All authors contributed to the study conception and design. The first draft of the manuscript was written by Sahar Shokouhirad and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability The datasets generated and presented in this study are available on reasonable request from the corresponding author.