



Seismic Zone Map for India Based on Cluster Analysis of Uniform Hazard Response Spectra

BHARGAVI PODILI¹ and S. T. G. RAGHUKANTH¹

Abstract—A novel methodology for obtaining a seismic zone map of India is demonstrated in this study, wherein a concrete theoretical framework is provided for deriving the zones and the respective zonal response spectra. The method involves time series clustering of uniform hazard response spectra (UHRS) that were obtained for the entire country on a $0.1^\circ \times 0.1^\circ$ grid by performing probabilistic analysis corresponding to a 2475-year return period. The Euclidean distance between the UHRS values at all periods (27 data points between 0.01 s and 5 s) was taken as the similarity measure in an evolutionary particle swarm optimization algorithm. The analysis was conducted with a swarm population of 100 over 3000 iterations, and the mean UHRS of the resulting clusters was assumed as the cluster centre. Various quality/validity indices including the compactness measure, similarity measure, combined measure and Dunn Index were used to verify the results of the clustering. Based on these clusters, the entire country can be divided into seven zones, with a unique zonal spectrum for each zone.

Keywords: Seismic zones, India zone map, probabilistic seismic hazard analysis, uniform hazard response spectra, cluster analysis.

1. Introduction

The Indian subcontinent and its adjoining regions have experienced many devastating earthquakes over the past century. The damage caused by the 2005 Kashmir earthquake and the recent 2015 Nepal earthquake highlighted the high seismic risk in the Himalayan seismogenic zone. In addition, the eastern and western boundaries of the subcontinent are frequently prone to earthquakes resulting from subduction plate boundaries. In this regard, even though peninsular India has lower seismic frequency, it is nevertheless prone to damaging events such as

the 1967 Koyna earthquake (M_w 6.5), the 1993 Latur earthquake (M_w 6.3) and the 2001 Bhuj earthquake (M_w 7.7) (NDMA, 2022). Therefore, appropriate hazard-based seismic design is imperative in mitigating imminent seismic risk. Accordingly, appropriate design values that integrate the seismic hazard are required for a meticulous earthquake-resistant design. Indeed, most of the international seismic design codes, such as the American Society of Civil Engineers ASCE-7 of the USA, the Building Standard Law (BSL) of Japan and EUROCODE 8 of Europe, recommend site-specific seismic hazard analysis for obtaining earthquake-resistant design values for complicated structures. After performing any hazard analysis, one would typically represent the results as contour maps of the ground motion parameter, usually that of peak ground acceleration (PGA) or response spectral acceleration (S_a) at a particular period of interest ($T = 0.1$ s, 2 s). However, most of the seismic design codes across the world (EUROCODE 8, IS 1893–2016) have provided zone maps instead of contour maps. The main reason for choosing a simple zone map over a contour map is to provide the user a simplistic solution for obtaining the design values. Therefore, an appropriate seismic zone map is necessary for better earthquake-resistant design practices, especially in developing countries such as India, where a larger demographic is confined to non-urban areas where strict design principles are not often followed.

The Indian Standard code IS 1893 (2016) provides basic guidelines for earthquake-resistant design of structures in India. In this regard, section 6.4 of the code specifies a design horizontal seismic coefficient, which is dependent on zone factor (Z), average acceleration response coefficient (S_a/g), importance factor (I) and response reduction factor (R). The latter two variables depend on the type of engineering

¹ Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, India. E-mail: bhargavi19892@gmail.com

structure, whereas the former two variables depend on probable local seismic forces. In order to provide location-based information on seismic forces, the IS code classified the entire country into four zones (II,

III, IV and V), as shown in Fig. 1. Consequently, zone factor (Z) is specified for each of these zones, along with a standard period dependent (between 0 and 4 s) design spectrum. Since the design spectrum

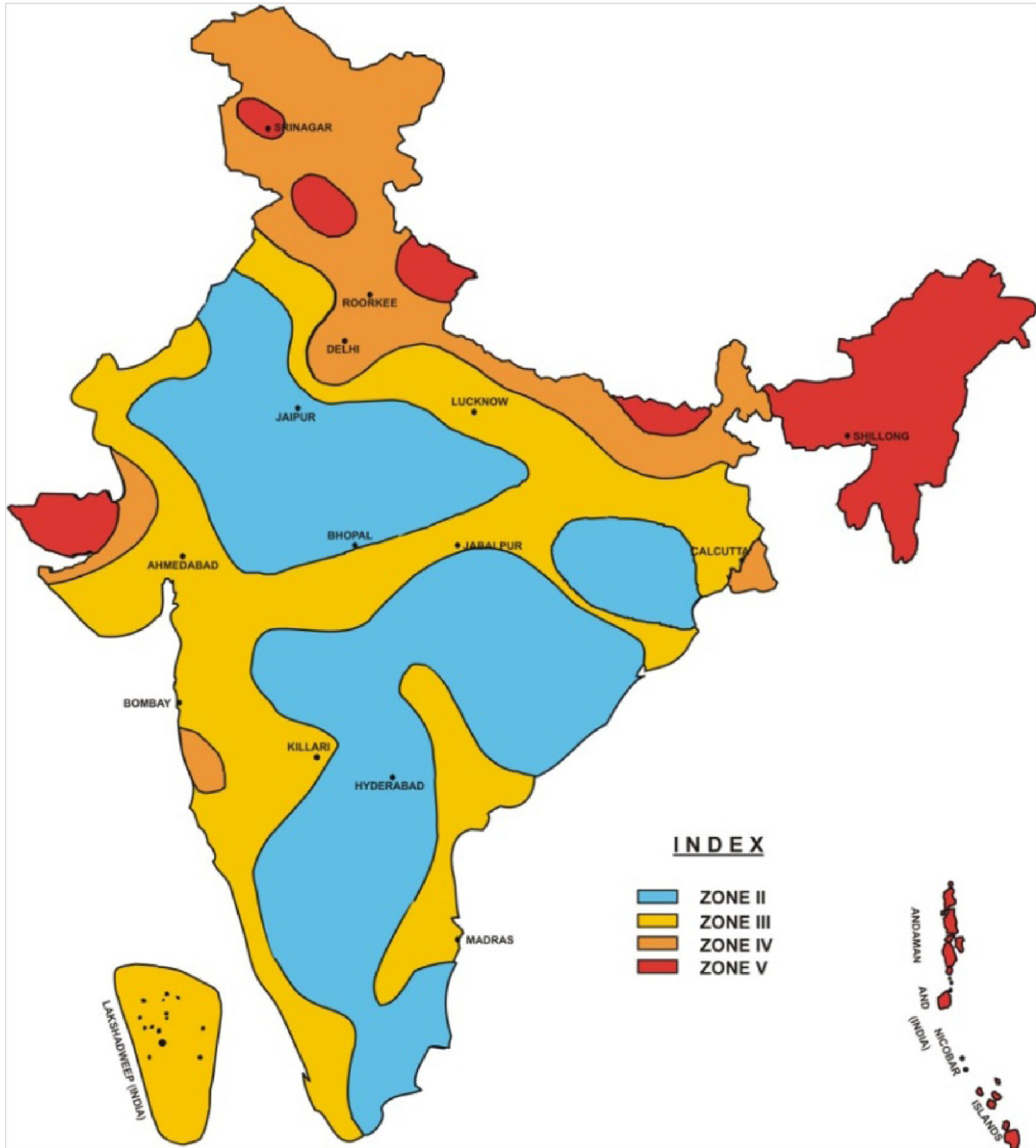


Figure 1
Seismic zone map of India, as given by IS 1893–2016

is normalized at $T = 0$ s, the zone factors provided in the code (0.1, 0.16, 0.24 and 0.36, respectively) specify the PGAs of each zone, which were supposed to be resulting from a maximum considered earthquake in the respective zone. It should be noted that the design spectral shape specified by the code remains the same, but its amplitudes vary depending on the zone factor. However, the current seismic zone map, which is supposed to be representative of varying seismic forces across the country, was based on damages observed during many of the past earthquakes, without any quantitative framework. This seismic zone map has been updated by the Bureau of Indian Standards (BIS) since 1962, when it was first proposed on the basis of the isoseismic map developed by the Geological Survey of India (GSI) in 1935 (BIS, 1962). Moreover, neither the theoretical background for the zone factors nor the return period for the specified design spectra is mentioned in the code. Therefore, there is a need to verify and establish the seismic zones of the country and a zone-specific design spectrum should be provided, using a solid theoretical framework such as the earthquake hazard.

Many of the international building design codes such as the EUROCODE 8, ASCE-7 initially provided seismic zone maps based on the historic occurrence of damaging earthquakes. However, throughout the years, these design codes have been updated to represent the design values quantitatively. In fact, the current version of European code for seismic design—EUROCODE 8—has subdivided all the national territories into seismic zones based on the local hazard. This hazard was described in terms of a single parameter, i.e., the value of the reference PGA on type A ground corresponding to a reference return period of the seismic action (no-collapse requirement). Furthermore, the ASCE-07 and the UBC (Uniform Building Code) have evolved from seismic zone maps to separate spectral acceleration maps for both short (0.2 s) and long (1 s) periods corresponding to a 2475-year mean return period. On the other hand, there have been many studies in the past on seismic zonation and hazard studies in India as well. Early zonation maps proposed by Tandon (1956), Krishna (1959), Guha (1962) and Gubin (1971), were developed on the basis of qualitative

measures such as the Modified Mercalli Intensity (MMI) intensity scale (Mohapatra & Mohanty, 2010). Since the deterministic analysis does not account for uncertainties associated with the magnitude and location of the source, probabilistic analysis is preferred to assess the hazard at a site. Basu and Nigam (1977) provided the first seismic zoning map for India (aside from BIS), based on a 100-year return period, following probabilistic seismic hazard analysis (PSHA). Khattri et al. (1984) made use of PSHA and provided a nationwide seismic zoning, with 24 zones based on a 10% probability of exceedance in 50 years. Bhatia et al. (1999) gave an 86-seismic-source-zone map based on major tectonic features and seismic trends for PSHA. Parvez et al. (2003) provided 15 regional polygons based on structural models, seismogenic zones, Q structure, focal mechanism and earthquake catalogue. Later, the National Disaster Management Authority (NDMA) divided the country into 32 seismogenic zones (NDMA, 2010). The final basis for zoning of most of the above studies is the design PGA value. However, none of these studies have made use of the complete response spectra across all the usable periods, to derive a zoning map. Even the International Building Code (IBC, 2015) recommends scaling with respect to periods of 0.2 s and 1 s, thereby providing both short-period and long-period design spectra. If such an attempt has been made and a response spectrum is given for each zone, one could obtain a zone-specific spectrum, which would be useful in earthquake-resistant design of all types of structures, especially in high-risk zones. This would in turn provide an engineer with site-specific design spectra for the entire country. Therefore, based on the above literature, the following objectives have been identified.

1. Derive a zone map based on grouping of the hazard-consistent-response spectra on the basis of pattern recognition
2. Develop design spectra for each of the resulting zones

The main principle adopted in achieving the above objectives is that the overall hazard in each seismic zone should be constant. In order to obtain the key component for the current study, i.e., the hazard-consistent response spectra for the entire

country, PSHA is carried out to obtain uniform hazard response spectra (UHRS), as given in Sreejaya et al. (2022), where probabilistic analysis was conducted, accommodating many of the recent advancements in PSHA and a comprehensive earthquake catalogue. An appropriate clustering technique is then adopted to derive a zone map, analogous to that of the seismic zone map provided by the IS 1893 (2016).

2. Uniform Hazard Response Spectra

This section provides a brief review of the procedure involved in deriving uniform hazard response spectra for India. The current study is performed in tandem with Sreejaya et al. (2022), in which PSHA of the entire Indian subcontinent was carried out to develop hazard maps. The basic methodology of PSHA is usually carried out in four phases: (1) identifying and defining seismic sources; (2) determining recurring characteristics of these seismic sources; (3) identifying or deriving ground motion prediction models; and (4) deriving hazard in terms of exceedance probabilities (Yu et al., 2011).

Seismic sources over the entire country were identified in the form of active faults and epicentres of past seismicity. The fault map of India, which is based on the seismo-tectonic map provided by the GSI (GSI, 2000) constitutes 1838 faults in total (Fig. 2). Detailed explanation regarding the seismo-tectonic setting of India was provided in Sreejaya et al. (2022). Further, an updated earthquake catalogue based on the NDMA (2010) register was considered, which consists of 27,869 independent earthquakes [declustered out of 68,016 events, based on the Gardner–Knopoff (1974)–Uhrhammer (1986) approach] that were registered between 2600 BC and December 2019. To derive recurring characteristics of these seismic sources, the 32 seismogenic zones of India (Fig. 2), as defined by NDMA (2010) are considered. Consequently, the recurrence characteristics of these previously identified sources were quantified zone-wise using the truncated Gutenberg–Richter (G-R) relation, wherein the controlling seismicity parameters include the upper bound of source magnitude (M_{\max}) and recurrence parameter (b). The

latter parameter was derived using two popular maximum likelihood methods proposed by Kijko and Graham (1998) and Weichert (1980). After obtaining the mean annual rate of exceedance with respect to the source magnitude (M_w) using the truncated G-R relation, the seismic activity rate was obtained using an elliptical smoothed seismicity model (Lapajne et al., 2003) based on the earthquake catalogue compiled for this purpose. Based on these parameters, the probability density function for M_w was obtained.

In addition, the probability density function for the intensity parameter was derived using several local and global ground motion prediction equations (GMPEs) applicable for the region, using a logic tree approach, to account for epistemic uncertainties. The weights of this logic tree were obtained using the normalized ranking method proposed by Kale (2019). Since the Indian subcontinent constitutes different tectonics, such as the active and subduction regions in the north and stable continental region in the south, appropriate GMPEs were selected. Table 1 provides a list of all the GMPEs used in the analysis.

The probabilistic analysis was then carried out for NEHRP (National Earthquake Hazards Reduction Program) B-C type site classes ($V_{s30} = 760$ m/s). It should be noted here that the NEHRP B-C site class is equivalent to IS 1893–2016 Type 1 site class. After obtaining all the necessary probability density functions, the exceedance probability of S_a for a particular return period was calculated. The probability of a ground motion intensity exceeding a particular level was calculated based on the probability of that intensity level occurring at a site, for a magnitude distance combination ($P[Y > y^* | R_i, M_w]$) and the probability of occurrence of the same magnitude event within the vicinity of the site.

$$P(Y > y^* \text{ in } T \text{ years}) = 1 - e^{-\mu_{y^*} T} \quad (1)$$

$$\mu_{y^*} = \sum_{i=1}^K \tilde{n}_i(m_0) \int_{M_w=m_0}^{m_{\max}} P[Y > y^* | R_i, M_w] f_{Mi}(M_w) dM_w \quad (2)$$

Here, μ_{y^*} is the mean annual rate of exceedance of ground motion parameter y^* ; $\tilde{n}_i(m_0)$ is the elliptically smoothed seismic activity rate; $P[Y > y^* | R_i, M_w]$ is the conditional probability that the parameter y^*

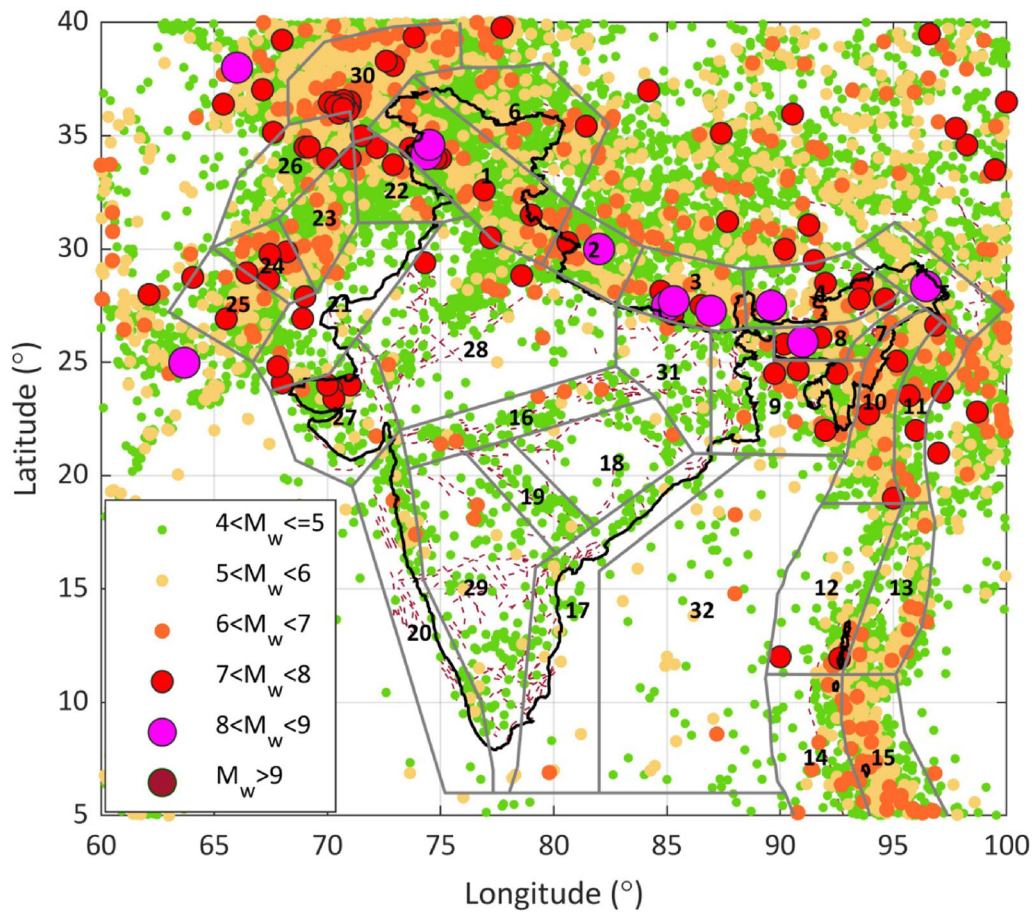


Figure 2

Seismo-tectonic map of India, showing past seismicity along with the 32 seismogenic zones over which UHRS was derived

would be exceeded for distance R_i and magnitude M_w ; and $f_{Mi}(M_w)$ is the probability density function for source magnitude. The mean annual rate of exceedance and the resulting uniform hazard response spectra (UHRS) curves for the entire country on a $0.1^\circ \times 0.1^\circ$ grid were obtained by Sreejaya et al. (2022). The resulting hazard maps were derived for PGA and 5% damped S_a at 0.2 s and 1 s, for various return periods (73, 475, 2475, 4975 and 9975 years). It was also reported that the site-specific PGA values thus obtained from the above analysis are almost twice those of the zone factors specified by the IS code, for a few cities in the Himalayas and North-East India.

The key input of the current study, i.e., hazard-consistent response spectra, is obtained after the

above analysis. The UHRS, which constitutes 5% damped S_a values obtained at 27 periods between 0.01 s and 5 s, for a 2475-year return period, i.e., 2% probability of exceedance in 50 years, is selected as the appropriate input for zoning. Figure 3 shows the contour maps derived at periods of 0.01 s, 0.2 s, 1.0 s and 5.0 s. A total of 14,911 UHRS curves are obtained for the analysis, which are shown in Fig. 4 along with the mean and confidence intervals. In order to understand the behaviour of UHRS with the seismogenic zones of the country, various properties of the curve including the initial, peak and end values, along with the piecewise slopes between these points, are investigated. These preliminary investigations across different important cities indicate that the long-period slope of the UHRS changes for many

Table 1
List of GMPEs for each of the four different regions within India

	Region	Type	GMPE used
01	Himalayas: active tectonic region	Local	GT 2018 (Gupta & Trifunac, 2018) DR 2020 (Dhanya & Raghukanth, 2020)
		Global	BBSA 2014 (Boore et al., 2014) Ask 2014 (Abrahamson et al., 2014) CB 2014 (Campbell & Bozorgnia, 2014) CY 2014 (Chiou & Youngs, 2014)
02	Indo-Gangetic plains: active tectonic region with sedimentary basin	Local	SSSA 2017 (Singh et al., 2017)
		Global	BSSA 2014 ASK 2014 CB 2014 CY 2014
03	Peninsular India: stable continental region	Global	NGA East (Goulet et al., 2021) AB 2006 (Atkinson & Boore, 2006)
04	North-East India: subduction zone	Global	BC Hydro 2016 (Abrahamson et al., 2016) Zhao 2006 (Zhao et al., 2006) Kanno 2006 (Kanno et al., 2006) AB 2003 (Atkinson & Boore, 2003) Youngs 1997 (Youngs et al., 1997)

of these cities, suggesting a steeper curve for cities in high-hazard zones. Moreover, the asymptotic PGA and the maximum/peak value vary for different regions as well. Thus, the entire UHRS across all periods would help to identify different zones according to their probable seismicity.

3. Cluster Analysis

In order to obtain a zoning map based on the entire UHRS across all the periods, time series grouping techniques such as the clustering analysis should be performed, following a proper protocol. There are four key components that should be established prior to performing a time series clustering: (1) data representation, i.e., how the data to be analysed are represented—whether as raw data or modified data after dimensionality reduction (since pure representation of the data would usually lead to higher computational costs); (2) similarity measure, which would serve as the basis for obtaining groups or clusters; (3) cluster prototype, i.e., obtaining a cluster centre for each group by choosing an appropriate statistical central tendency technique; and (4) the algorithm that is to be applied in the clustering method (Kamalzadeh et al., 2017).

The similarity measure should be selected based on the properties or features of the time series that are considered as the basis for grouping. It should be noted that all previous zoning maps of several international codes have been obtained on the basis of design PGA values on hard rock strata. In such cases, a simple Euclidean distance measure between the design PGA value of each data point would give effective results. On the other hand, since the entire UHRS is considered for clustering in the current case, and since there are only 27 data points in each UHRS, the Euclidean distance between all the corresponding data points simultaneously is treated as the distance measure. Subsequently, since the data points are limited and the computational costs are minimal, no data reduction techniques are applied and the entire raw dataset is used for analysis. Moreover, since all the time series used for clustering, i.e., the UHRS are of equal length, a simple method such as the average sequencing which includes obtaining the mean of all the UHRS at each period, is used to obtain the cluster prototype.

There are two types of clustering algorithms—evolutionary and non-evolutionary—wherein the non-evolutionary algorithms usually depend heavily on the initial solution and are deemed inadequate for dynamic or time series clustering as opposed to

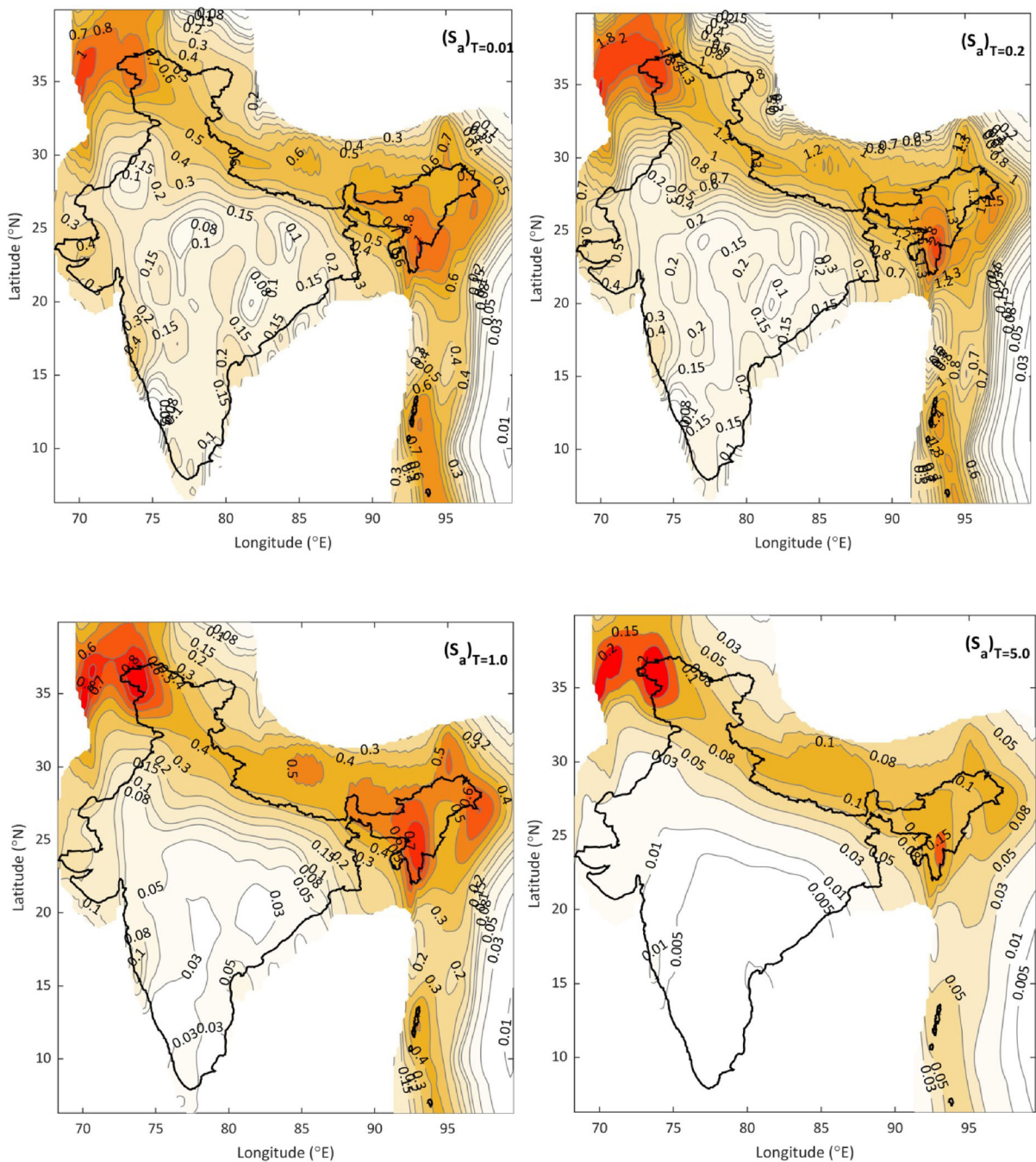


Figure 3

Contour maps showing the results of PSHA for a 2475-year return period UHRS at periods $T = 0.01$ s, 0.2 s, 1.0 s and 5.0 s

traditional static clustering techniques. On the other hand, evolutionary algorithms, which are usually designed based on the natural selection or collective intelligence exhibited by living beings, are

considered appropriate for the current case. These evolutionary algorithms include genetic optimization, ant-based clustering and particle swarm optimization (PSO). Among the available options, the PSO

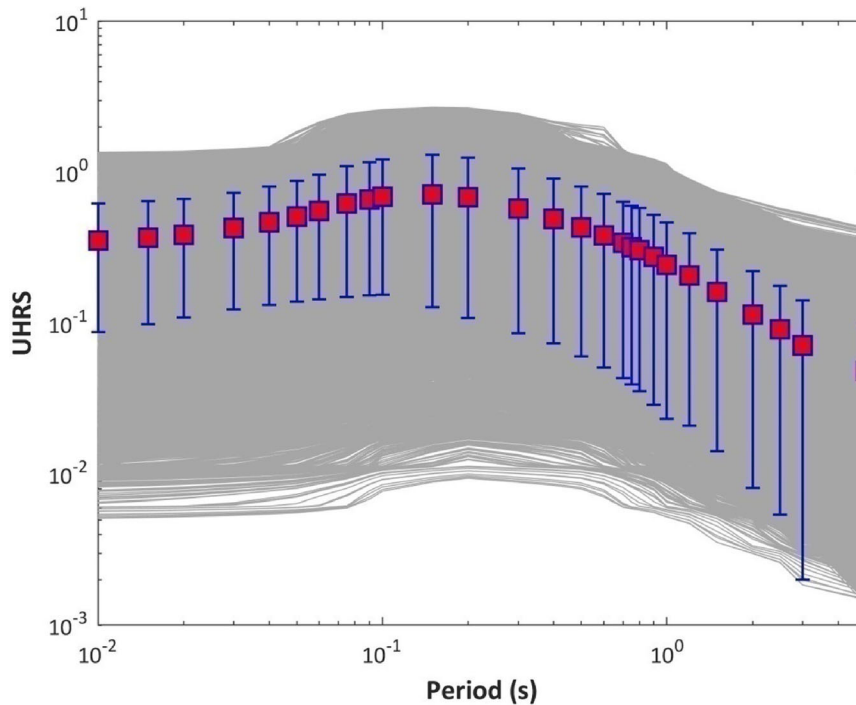


Figure 4

UHRS data used for obtaining clusters, showing the mean value and 5–95% confidence interval across all periods

algorithm is selected to perform the clustering analysis, due to its efficiency. The entire PSO algorithm in detail is given in Appendix A. In the current case, since the objective function involves clustering of the given time series by comparing the Euclidean distance between the respective 27-period UHRS, at each iteration, all the input time series (UHRS) are swarming towards their respective cluster prototypes, thus forming the final groupings. Here, the cluster prototypes are the cluster centres of the respective groups obtained from the mean of the input UHRS. Following the recommendations of the optimum population size (Piotrowski et al., 2020), the swarm size is taken as 100 and the iterations are run for $n = 3000$. Further, the PSO parameters (as described in Appendix A) such as the inertia weight w is taken as 1, while the cognitive and social acceleration parameters c_1 and c_2 are taken as 1.5 and 2.0, respectively. It should be noted that the clustering thus performed can be categorized as unsupervised clustering. Even though a cluster number is taken as an input in the above algorithm, the PSO technique

ensures that each input time series (UHRS) is flocked towards the most appropriate cluster centre, thereby resulting in fewer clusters than the input.

4. Validation of Cluster Analysis

Visual inspection of the clustering centres along with samples in the respective clusters would provide

Table 2

Validity measures obtained for different cases with varying numbers of clusters

Number of clusters (K)	Compactness measure (F_{CM})	Separation measure (F_{SM})	Combined measure ($F_{Combined}$)	Dunn Index (DI)
2	0.7916	0.5000	0.4041	0.00068
3	0.5768	0.1667	0.3537	0.0009
4	0.5162	0.0833	0.3364	0.0010
5	0.5029	0.0500	0.3370	0.0011
6	0.4384	0.0330	0.2970	0.0011
7	0.4149	0.0238	0.2833	0.0011

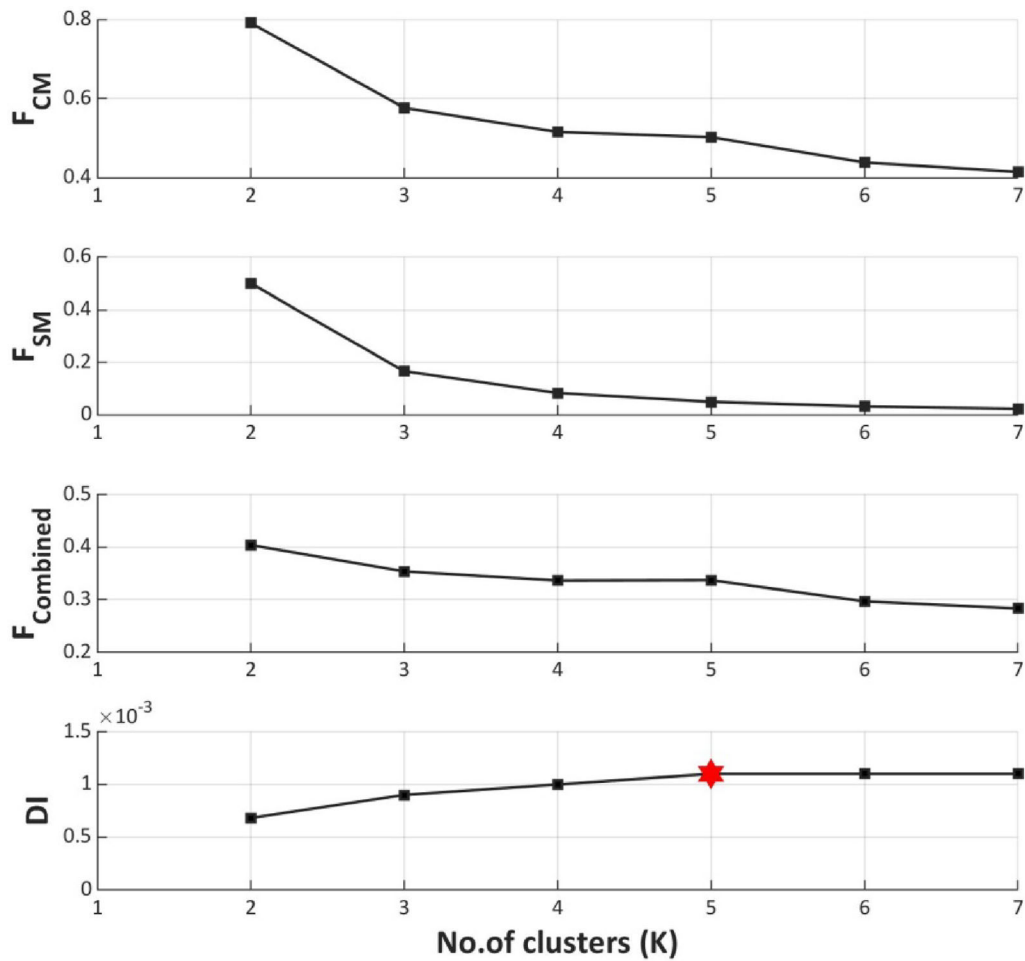


Figure 5

Comparison of the quality measures obtained for results with different numbers of clusters (K): Dunn Index (DI), compactness (F_{CM}), separation (F_{SM}) and combined ($F_{combined}$) measures. Since all four indices begin saturating at five clusters (red star), the entire UHRS can be grouped into five clusters

a fair idea regarding the goodness of clusters. However, in order to obtain quantitative validation on the optimal number of clusters, one has to make use of validity measures, which are represented using mathematical formulations. Even though many external validity indices such as the Dunn Index, Xie–Beni index etc. are available, since the current study involves unsupervised clustering, internal indices (Aghabozorgi et al., 2015) such as the compactness and separation measures are deemed more appropriate to validate results of the above cluster analysis. The compactness measure (F_{CM}) gives a scalar representation of the similarity of the samples

in a cluster, which can be calculated in terms of ‘intra-cluster distance’, as given in Eq. 3. On the other hand, the Separation Measure (F_{SM}) represents the distance between clusters, which can also be referred to as ‘inter-cluster distance’ (Ahmadi et al., 2010).

$$F_{CM}(\text{cc}^{(1)}, \dots, \text{cc}^{(K)}) = \frac{1}{K} \sum_{k=1}^K \frac{1}{n_k} \sum_{i=1}^{n_k} d(\text{cc}^{(k)}, s_i^{(k)}) \tag{3}$$

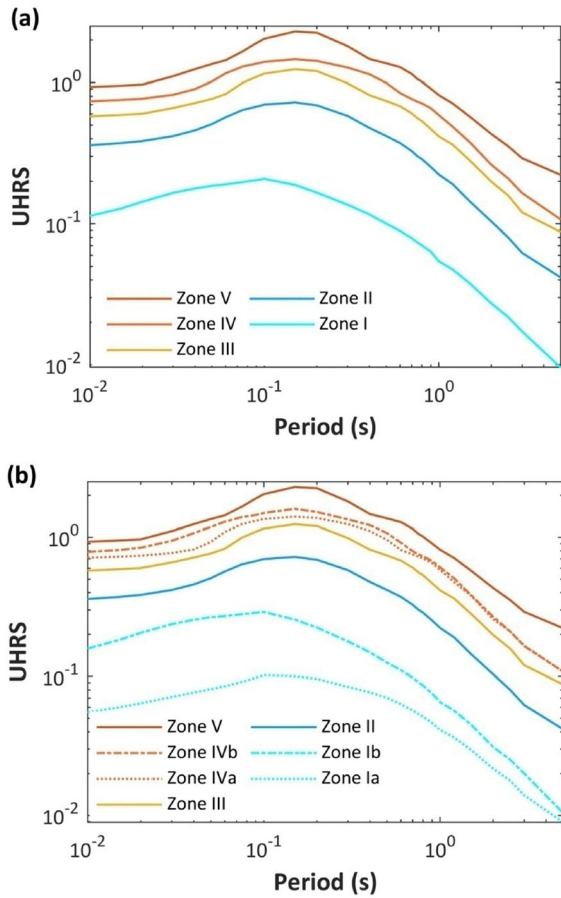


Figure 6

Mean UHRS curves for each of the five zones (a) and the respective sub-zones of zone IV and zone I (b). It can be seen that the mean UHRS curves of the sub-zones are clearly not distinct at long periods

$$F_{SM}(cc^{(1)}, \dots, cc^{(K)}) = \frac{1}{K(K-1)} \sum_{j=1}^K \sum_{k=j+1}^K d(cc^{(j)}, cc^{(k)}) \quad (4)$$

Here, F_{CM} and F_{SM} are compactness and separation measures; cc represents the cluster centre of the respective cluster; K is the total number of clusters; s_i represents samples or time series of each cluster; $d(cc^{(k)}, s_i^{(k)})$ represents the distance measure between the cluster centre and i th sample of the respective cluster k ; $d(cc^{(j)}, cc^{(k)})$ represents the distance measure between cluster centres of j th and k th clusters.

$$F_{combined} = w_1 F_{CM} - w_2 F_{SM} \quad (5)$$

Table 3

Period-dependent uniform hazard response spectra proposed for each seismic zone

Period (s)	Zone I		Zone II	Zone III	Zone IV		Zone V
	a	b			a	b	
0.01	0.0553	0.1582	0.36	0.5767	0.7139	0.7864	0.9306
0.015	0.0598	0.1822	0.3726	0.5898	0.7264	0.8107	0.9488
0.02	0.064	0.2055	0.3853	0.6021	0.7379	0.8429	0.9673
0.03	0.0711	0.2372	0.418	0.6605	0.7721	0.9442	1.1082
0.04	0.0765	0.2554	0.4585	0.7159	0.8179	1.0713	1.2425
0.05	0.0809	0.266	0.5085	0.7661	0.9276	1.1879	1.3511
0.06	0.0849	0.2714	0.5683	0.8289	1.0906	1.2929	1.4377
0.075	0.0912	0.2798	0.6397	0.9947	1.2429	1.3964	1.6579
0.09	0.0976	0.2863	0.6752	1.0976	1.3139	1.4439	1.9009
0.1	0.1021	0.2908	0.6985	1.1554	1.3551	1.4941	2.0406
0.15	0.1002	0.2556	0.7235	1.2416	1.4043	1.5987	2.2977
0.2	0.0958	0.2245	0.6902	1.2046	1.3771	1.5176	2.2507
0.3	0.084	0.1781	0.5809	0.9878	1.2425	1.3442	1.8082
0.4	0.0767	0.1486	0.4776	0.8156	1.1098	1.2221	1.4664
0.5	0.0701	0.1253	0.418	0.7405	0.9585	1.072	1.367
0.6	0.0632	0.1107	0.3728	0.6802	0.8104	0.9146	1.2841
0.7	0.0568	0.0976	0.3267	0.605	0.7537	0.7954	1.1486
0.75	0.0539	0.0911	0.3026	0.5625	0.7277	0.7556	1.0638
0.8	0.0514	0.086	0.287	0.5337	0.707	0.7317	1.013
0.9	0.0463	0.0756	0.2525	0.4687	0.6539	0.6764	0.9047
1	0.0412	0.0654	0.223	0.4158	0.5837	0.6095	0.8154
1.2	0.0364	0.0568	0.1908	0.3636	0.477	0.5087	0.7117
1.5	0.0296	0.045	0.1445	0.2817	0.381	0.384	0.5759
2	0.0217	0.031	0.1036	0.1986	0.2576	0.2686	0.4325
2.5	0.0178	0.0252	0.08	0.1582	0.2102	0.2101	0.3538
3	0.014	0.02	0.0623	0.1202	0.1663	0.165	0.2912
5	0.0089	0.0104	0.0417	0.0874	0.1077	0.1078	0.2217

$$DI = \min_{1 \leq k \leq K} \left\{ \min_{k+1 \leq l \leq K} \left\{ \frac{\min_{x \in C^{(k)}, y \in C^{(l)}} d(x, y)}{\max_{1 \leq m \leq K} (\max_{x, y \in C^{(m)}} d(x, y))} \right\} \right\} \quad (6)$$

The objective of any good clustering technique is to minimize the cumulative F_{CM} and to maximize the cumulative F_{SM} . In order to achieve both these objectives, the values of the combined measure (as given in Eq. 5) are also noted. The weights w_1 and w_2 should be selected such that $w_1 + w_2 = 1$ (Ahmadi et al., 2010). In the current study, the clusters are sensitive to the number of iterations and population size (n, p). Moreover, it should also be noted that the current study is a special case in which, as the grouping or cluster number increases, the separation between the cluster centres decreases. Therefore, high

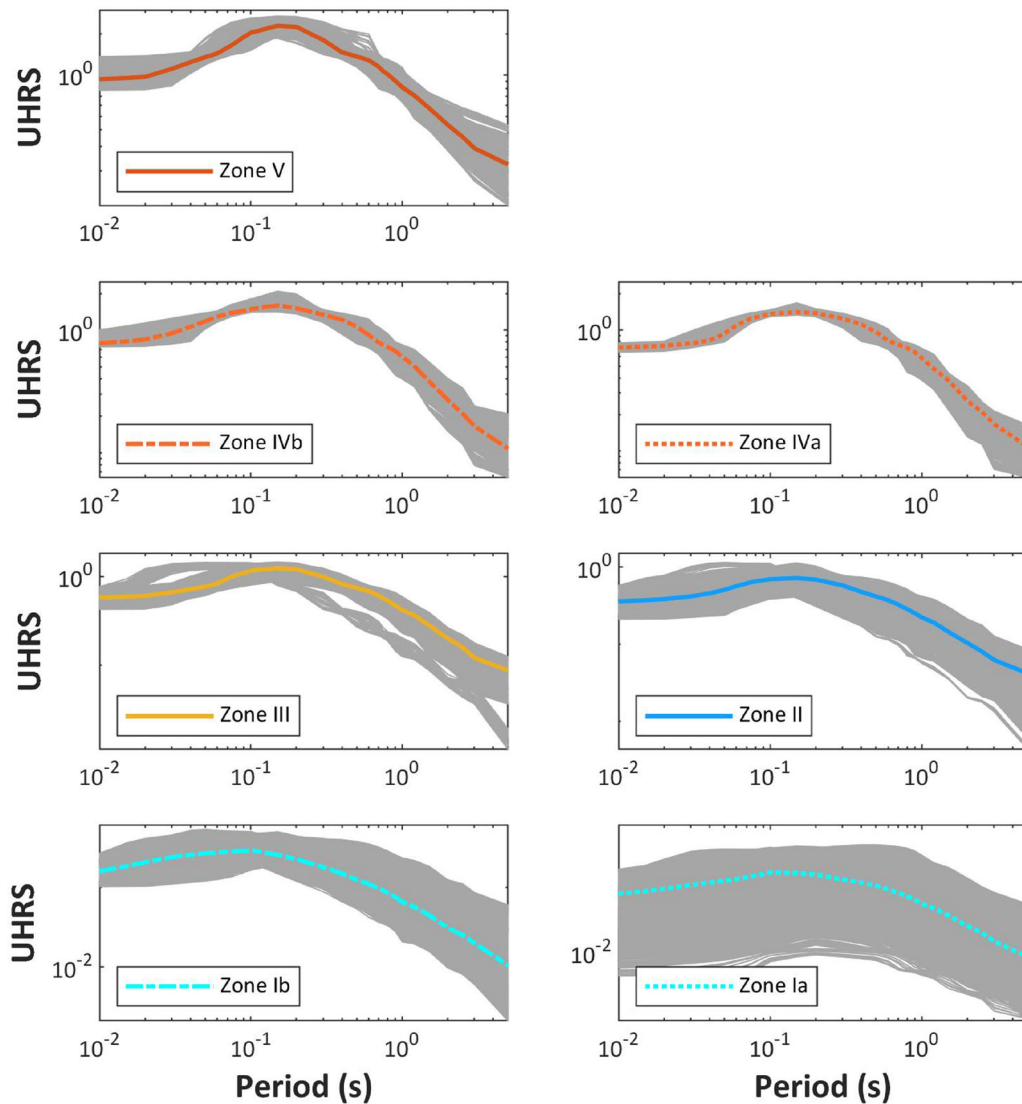


Figure 7

Cluster centres for each of the five main zones and the two sub-zones against the samples (response spectra) of each cluster

weight (70%) is given to F_{CM} , thereby selecting the values for w_1 and w_2 as 0.7 and 0.3, respectively. Further, the optimal number of clusters is selected using F_{CM} , F_{SM} , $F_{combined}$ and the Dunn Index (as given in Eq. 6), which is one of the standard validity measures (Dunn, 1973) devised for cluster analysis. Table 2 gives the values of all these validity measures for each of the clustering results obtained for different (n, p) combinations. The same results are visually represented in Fig. 5, wherein the UHRS is grouped into different numbers of clusters (1, 2, 3, 4, 5, 6 and

7) in each of the cases. It can be noted that the F_{CM} decreases with an increase in the number of clusters K , indicating that the compactness of the samples in each cluster increases as the data are clustered into a greater number of groups. On the other hand, F_{SM} decreases gradually with an increase in K value. The value of the combined measure $F_{combined}$, when given equal weight to F_{CM} and F_{SM} , is lowest for $K = 2$. However, the Dunn Index increases with the number of clusters until $K = 5$, after which it remains constant. Therefore, when all four measures are

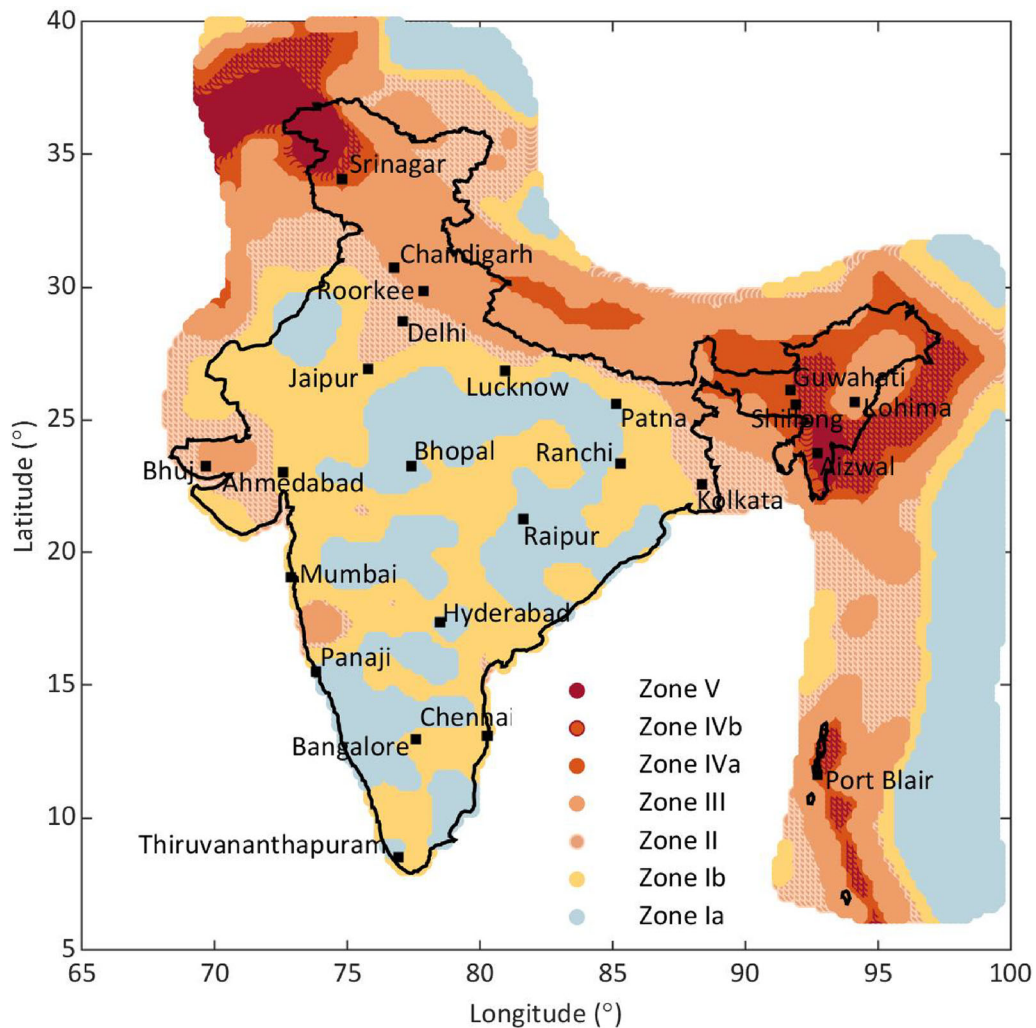


Figure 8

Proposed zone map of India showing a few of the metropolitan and urban agglomerations of the country

considered together, the plot indicates that the optimal number of groups for the current dataset is five.

5. Zone Maps Based on 2475-Year Return Period UHRS

Following the results of cluster analysis performed in the previous section, the entire grid of India is divided into five zones, based on the 27-period UHRS, which is obtained for a 2475-year return period. The cluster centres are assumed as the mean UHRS curves for each of these five zones. Figure 6a

shows all of these cluster centres. It was observed that the standard deviations with respect to the cluster centre are maximum for the second and fifth zones. Providing unusually high design values is not recommended unless it is absolutely necessary, as an engineer focuses equally on the economy of the project. Thus, all the resulting zones are further subdivided by performing the cluster analysis again on the data of each cluster. It was noted that only zone IV of the top three clusters yielded distinct UHRS curves after subdivision, albeit at short and intermediate periods, thus resulting in sub-zones IVb and IVa. The respective mean UHRS curves are

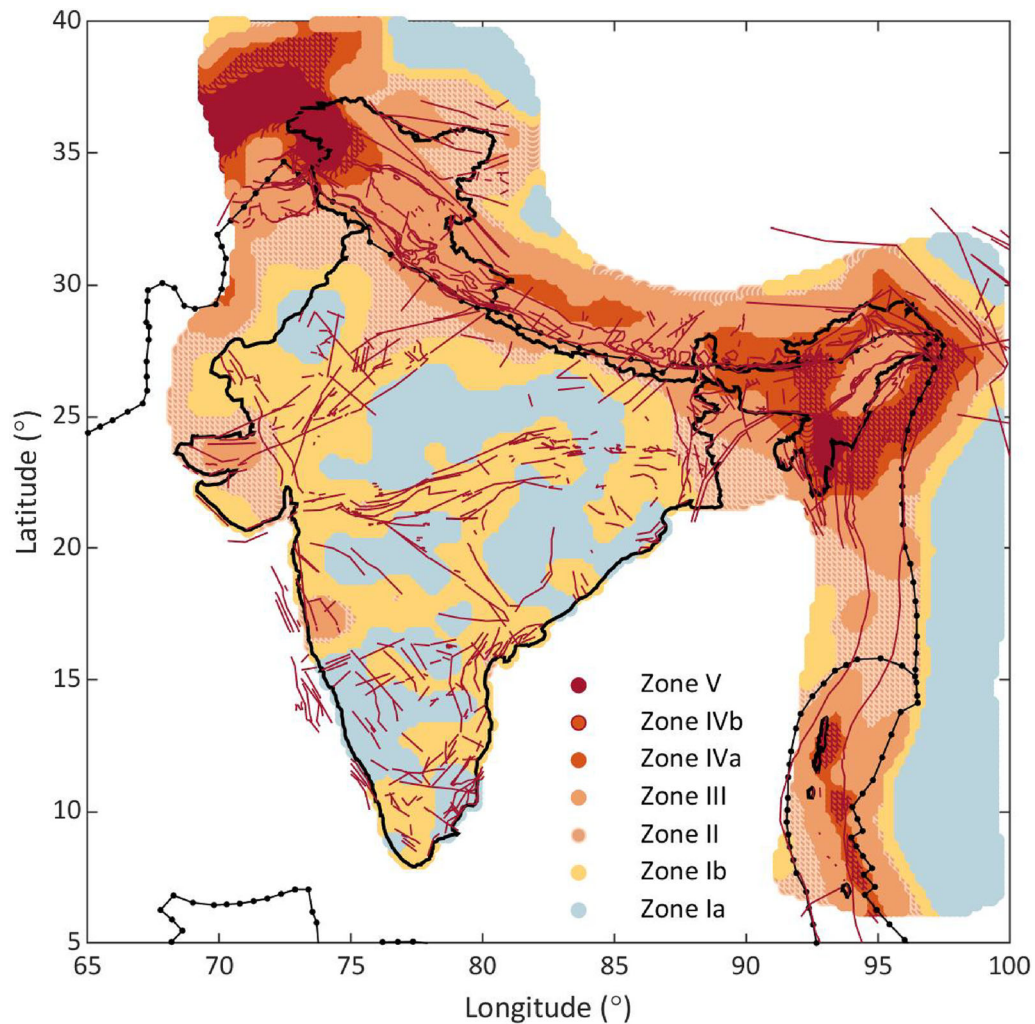


Figure 9

Proposed zone map of India against the active faults and tectonic boundaries of India

provided in Fig. 6b. Similarly, in order to avoid overestimations in low-hazard areas, a similar analysis is conducted for zone I where there is a large quantity of data below the mean UHRS curve. The resulting analysis provided sub-zones Ib and Ia (Fig. 6b). The mean UHRS curves are also provided in Table 3, for all the five zones and two sub-divisions. In addition, Fig. 7 shows all of these cluster centres against the input UHRS data, further illustrating the goodness of the analysis.

The resulting zone map is illustrated in Fig. 8, along with the important cities of India. In addition, the zone map is also plotted with the active faults of

the country in Fig. 9. It can be observed that zones V and IV, which are associated with high hazard values, contain the most active faults of the country. The same zone map is illustrated against the past seismicity of India in Fig. 10. It should be noted that the current zone map captures the hazard resulting from all the past devastating earthquakes such as the 1967 Koyna earthquake, the 2001 Bhuj earthquake, the 2005 Kashmir earthquake, and the 2015 Nepal earthquake, confining the epicentral regions of these events to zones V and IV. Therefore, the zone map resulting from the current analysis is effective in capturing the seismic hazard of India.

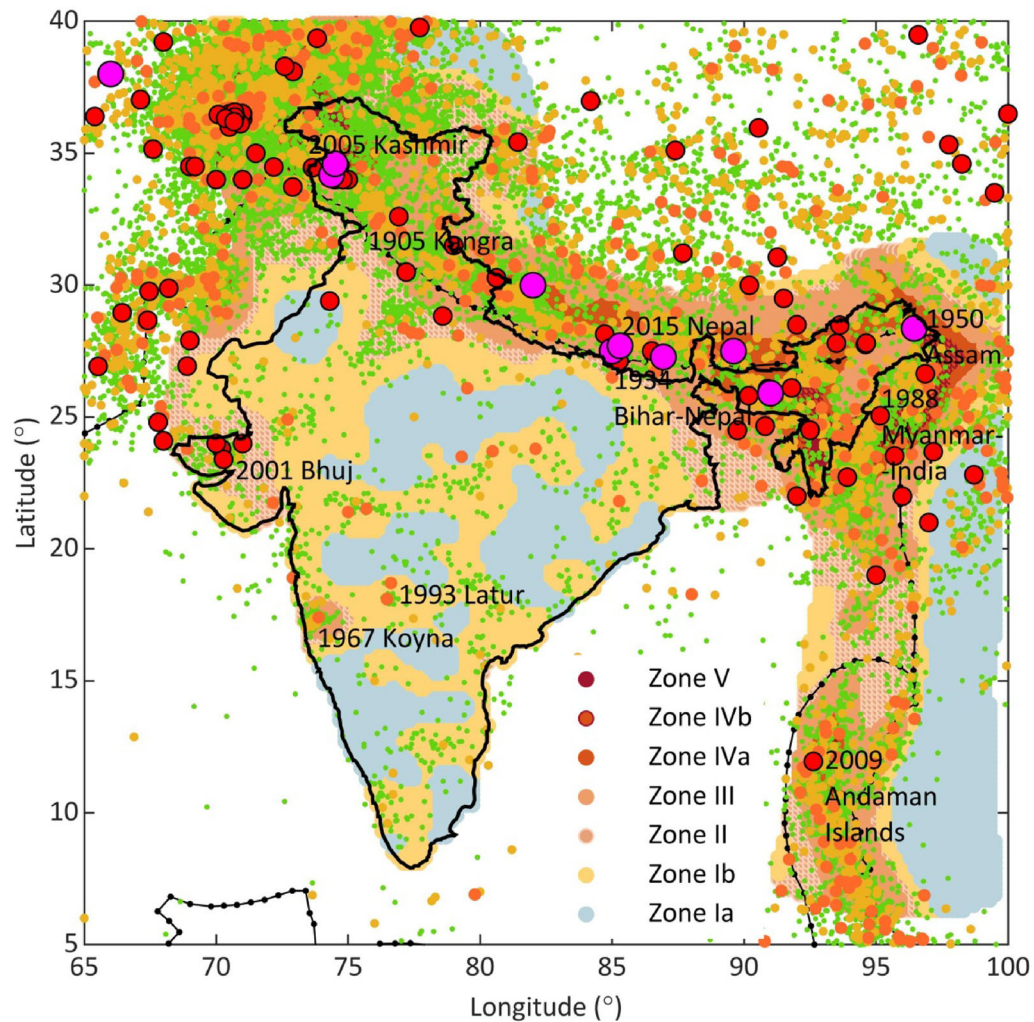


Figure 10

Proposed zone map of India against the past seismicity of India (magenta: $8 \leq M_w < 9$; red: $7 \leq M_w < 8$; orange: $6 \leq M_w < 7$; yellow: $5 \leq M_w < 6$; green: $4 \leq M_w < 5$)

6. Comparison with the Current Seismic Code IS 1893 (2016)

In order to compare with the design curve and the zone factors provided by the current seismic code of design IS 1893 (2016), the mean UHRS curves are normalized at $T = 0$ s. Figure 11 and Table 4 shows the respective normalized curves and the resulting zone factors. As represented in Table 4, zones V and IV of the current study are synonymous with zone V of IS 1893–2016. Further, zones III, II and I are synonymous with zones IV, III and II of the code,

respectively. The shape of the normalized spectra recommended by the IS code is conservative to that of each of the normalized spectra derived in this study. However, it should also be noted from the values of Table 4 that the current code severely underestimates the response spectra values, especially in high hazard zones. It should be noted that if a structural engineer disregards separate site specific analysis, the seismic code-provided values are directly used in structural design. The current code specifies a maximum design PGA of 0.36 g (zone V), while the hazard analysis specifies a design PGA of

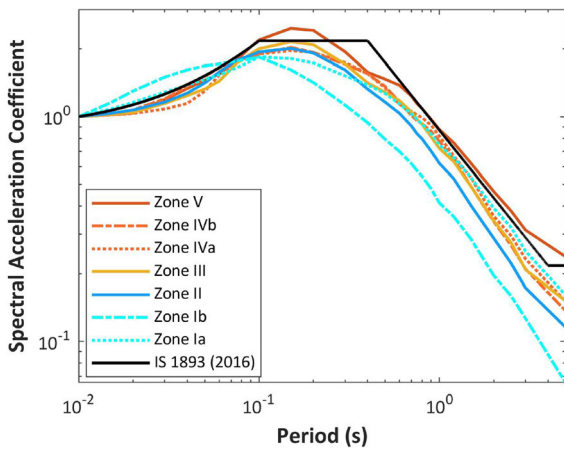


Figure 11

Comparison of normalized response spectra (with respect to $T = 0.01$ s) curves of each of the seismic zones obtained in this study against the normalized design curve specified by the IS 1893 (2016) for type I sites

Table 4

Comparison of zone factors provided in IS 1893 (2016) against the current study

IS 1893 (2016)		Current study	
Zone	Zone factor	Zone	Zone factor
		Ia	0.055
II	0.10	Ib	0.158
III	0.16	II	0.36
IV	0.24	III	0.576
V	0.36	IVa	0.713
		IVb	0.786
		V	0.93

0.93 g. Therefore, this large discrepancy in the code-given values would severely underestimate seismic hazard, especially in active regions of the country. Given the basis for the current seismic design code is observed intensity levels, there is an immediate need to verify and redefine the seismic design curves of the earthquake-resistant code of design for India, with a concrete theoretical approach such as the hazard analysis.

In addition, the zonal response spectra curves are compared against the actual UHRS curves calculated from hazard analysis and the recommendations provided by IS 1893 (2016). Figure 12a and b provide these comparisons for a few important cities of India

corresponding to the IS 1893–2016 zones IV and V, respectively. It can be observed that the cities in zone V of IS 1893 are categorized between zones IVb and IVa and III of the current study. In most of these cases, the zonal UHRS given in this study matches the UHRS obtained from probabilistic analysis. The only discrepancy is observed for Bhuj, which lies in the stable continental region of the country. This deviation is attributed to the different set of GMPEs (Table 1) used for Peninsular India to that of the active regions in the hazard analysis. Since the zonal UHRS is conservative to that of the hazard analysis curves, the zonal response spectrum given for zone V is validated. On the other hand, the IS 1893–2016 design curves severely underestimate the seismic demand for cities categorized under zones IVb and IVa in the current study. The comparison plots of zone III cities Kohima and Bhuj, whose zonal response spectra matches with that of IS 1893–2016 design curves, indicate that zone III of the current study is equivalent to zone V of the seismic code. The comparison plots of zone III cities Chandigarh and Roorkee in Fig. 12b, which show the underestimated code-based design curves, indicate that zone III of the current study is prone to high hazard levels comparable to that of zone IV of the seismic code.

Similarly, Fig. 13a and b provide these comparisons for a few important cities of India corresponding to the IS 1893–2016 zones III and II, respectively. As observed in the previous case, minor differences are observed between the zonal UHRS and hazard results for cities located in peninsular India. The zonal UHRS and IS 1893–2016 design curves are mostly complimenting each other without major deviations, for cities categorized under zone II in the current study. However, the comparison plots for cities in zone Ib and Ia of the current study indicate that the IS code-provided design curves are conservative to that of the zonal UHRS, wherein over-estimations are observed in very few cases. Overall, these comparison plots also indicate that the current seismic code provisions do not account for the high hazard levels observed in the active regions of India, especially those categorized under zones V, IVb and IVa in the current study.

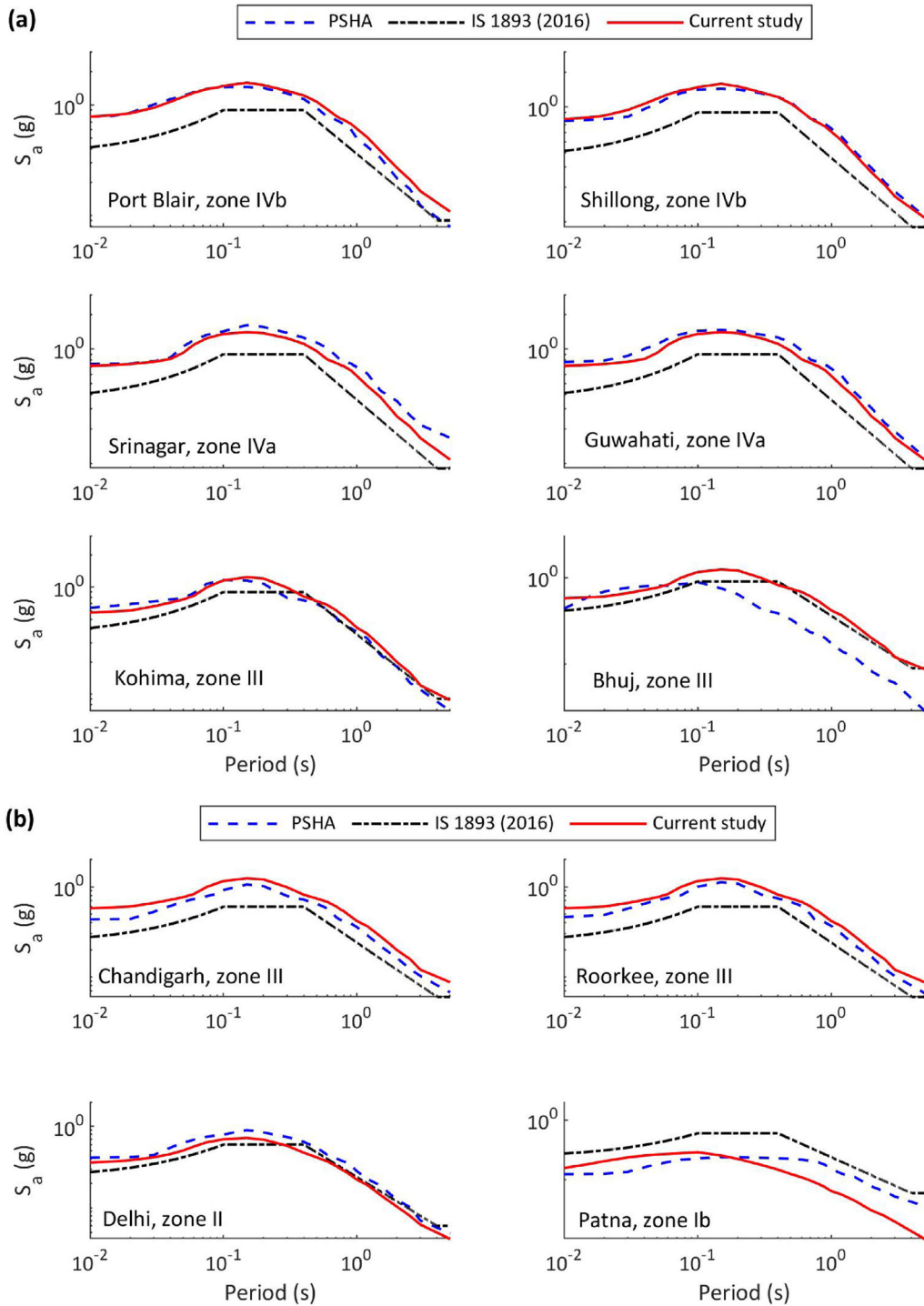


Figure 12

Comparison of response spectra curves proposed for each of the seismic zones (as given in fig) of the current study against the results of probabilistic seismic hazard analysis (PSHA) and the design curve specified by IS 1893 (2016) for **a** zone V and **b** zone IV

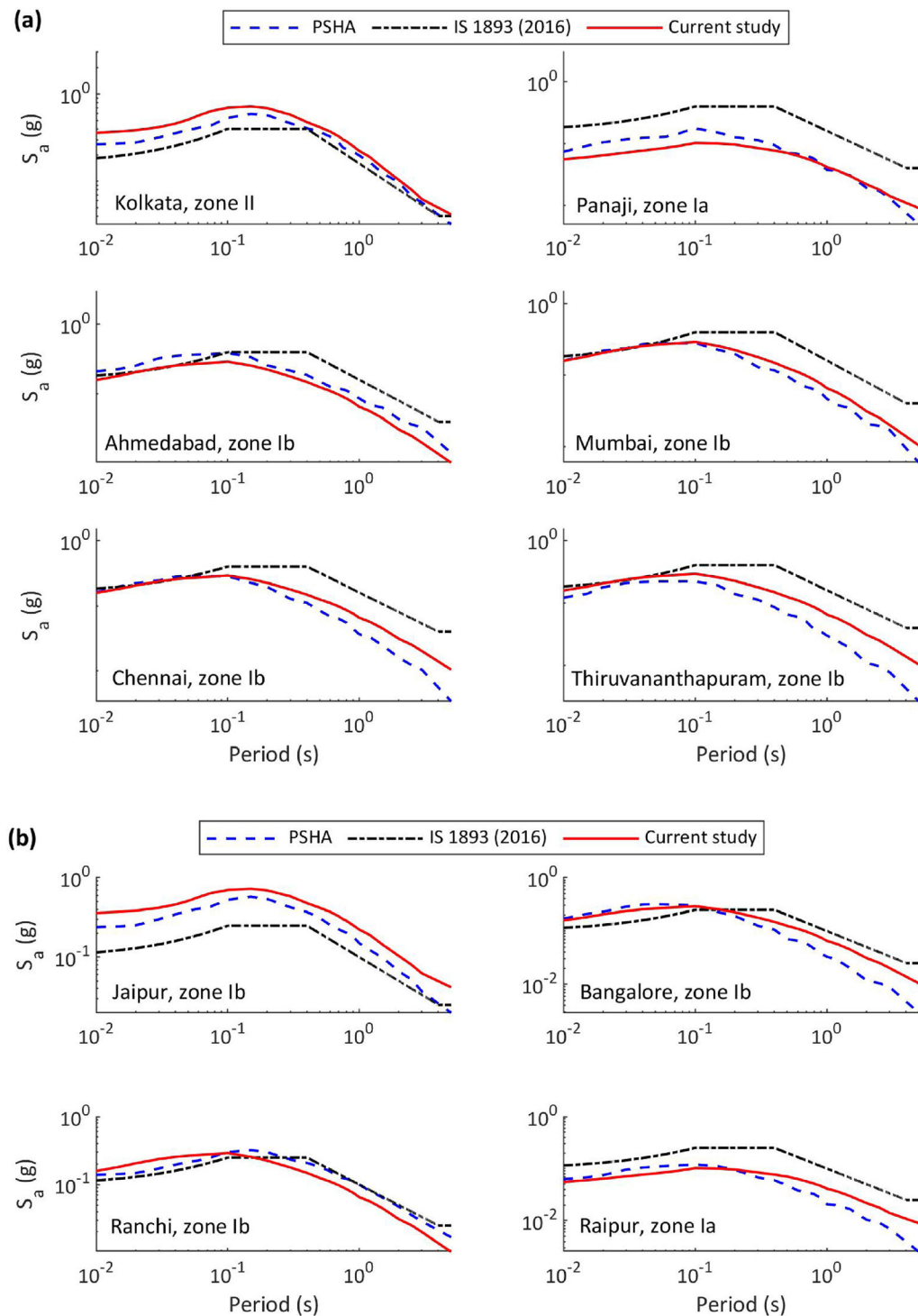


Figure 13

Comparison of response spectra curves proposed for each of the seismic zones (as given in fig) of the current study against the results of probabilistic seismic Hazard analysis (PSHA) and the design curve specified by IS 1893 (2016) for **a** zone III and **b** zone II

7. Summary and Conclusion

The preliminary analysis of earthquake-resistant design of structures involves quantifying the seismic forces of the region. Even though many important projects prefer site-specific hazard analysis, most engineers follow the seismic code of design IS 1893 (2016). However, the response spectra design curve specified by the code and the associated zone map was not obtained based on any concrete framework such as the hazard analysis. Moreover, all the previous traditional PSHA studies provided hazard maps based on PGA. In recent times, building codes such as the IBC and a few other studies (Sreejaya et al., 2022) have provided single short-period ($T = 0.2$ s) and long-period ($T = 1$ s) hazard maps. However, for an engineer to obtain a design curve for any type of structure at a particular site without resorting to site-specific analysis, a zone map based on the hazard at all periods, along with a mean hazard curve at all the respective periods, would be invaluable. The current study has provided this, while proposing a new approach for obtaining the seismic zone map. Here, cluster analysis was performed on the UHRS between periods of 0.0 s and 5 s, which was derived for India on a $0.1^\circ \times 0.1^\circ$ grid using probabilistic hazard analysis. The unsupervised statistical analysis resulted in a total of five clusters for the entire country. However, the problem was also studied from an engineering point of view wherein the economy of the design is also prioritized. Supervised clustering was then performed to reduce the high design values, thus resulting in a total of seven zones vis-à-vis seven clusters.

Even though many global seismic codes have derived zone maps based on 10% probability of exceedance in 50 years (475-year return period), since the NDMA of India adopted 2% probability of exceedance in 50 years as its basis, the current zone map is derived using a 2475-year return period UHRS. Since the design philosophy changes with the type of structure, corresponding zone factors and zonal response spectra can be calculated using the scaling factors that were derived between different return periods, as reported in NDMA (2022). It should also be noted that the current zone map and the corresponding zonal UHRS are derived using

results of the analysis that was carried out for Type I site class ($V_{s30} = 760$ m/s). Therefore, further analysis can be conducted to obtain zonal UHRS curves for different site conditions. Overall, the seismic zonation provided in the current study can also be used for investigations involving risk analysis, land planning and others, in addition to traditional seismic design of structures.

Author Contributions STG contributed to the conceptualization and supervision of this work, whereas concept development, data collection, analysis and preparation of the manuscript were performed by the corresponding author BP.

Funding

The authors declare that no funds, grants, or other support were received for this project or during the preparation of this manuscript. The authors have no relevant financial or non-financial interests to disclose.

Data Availability

The models developed in this study are given as electronic supplements. Moreover, the datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Appendix A

Particle Swarm Optimization

Particle swarm optimization (PSO), which is an evolutionary algorithm, makes use of the nature-bound swarm intelligence technique in solving an optimization problem (Kennedy & Eberhart, 1995). The logic in the algorithm is based on the natural behaviour or movement of birds/fishes in a swarm

(flock/school). In the algorithm, a set of particles with memory are considered to fill the solution space. The particles have an adaptable velocity, which determines their movement in the search space towards the solution.

The PSO algorithm starts with initialization of the position vectors (x_i) of all the particles in the swarm. These position vectors are assumed to be the solutions of the optimization problem. Each of these particles also has velocity vectors (v_i), which determines its respective position in the subsequent time step, as given by Eq. 2. At every iteration, the objective function or cost function is evaluated based on the particle's position x_i and the best position for the i th particle in the swarm is estimated. Since all the particles have memory, the current best position is compared against the personal best position (Pbest), which might have been observed in any of the previous iterations. Simultaneously, a global best position (Gbest) is obtained through comparison of all the particles' positions in the swarm. Accordingly, based on the Pbest and Gbest values, the velocity of the particle i is updated (Eq. 1) and the procedure is repeated for all the n iterations to obtain new positions for all the particles (Eq. 2). It should be noted that in each of these iterations, the position of the particle is updated towards the global best position, mimicking the swarm phenomena.

$$v_i^{n+1} = wv_i^n + c_1r_1^n(Pbest_i^n - x_i^n) + c_2r_2^n(Gbest_i^n - x_i^n) \quad (7)$$

$$x_i^{n+1} = x_i^n + v_i^{n+1} \quad (8)$$

Here, w is the inertia weight; c_1 is the cognitive acceleration parameter; c_2 is the social acceleration parameter; r_1 and r_2 are uniformly distributed random numbers bounded by [0 1] (Kennedy et al., 2001). The change in velocity and the bounds of the increased velocity at every time step is governed by the parameter w and the upper bound of the velocity vector (V_{max}). Consequently, the maximum step size of a particle in a single iteration is governed by parameters c_1 and c_2 , whereas r_1 and r_2 maintain the population diversity. The bounds for the velocity vector are obtained using Eqs. 3 and 4.

$$V_{max} = 0.1(\max(x) - \min(x)) \quad (9)$$

$$V_{min} = -V_{max} \quad (10)$$

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

REFERENCES

- Abrahamson, N., Gregor, N., & Addo, K. (2016). BC Hydro ground motion prediction equations for subduction earthquakes. *Earthquake Spectra*, 32(1), 23–44.
- Abrahamson, N. A., Silva, W. J., & Kamai, R. (2014). Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra*, 30(3), 1025–1055.
- Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering—A decade review. *Information Systems*, 53, 16–38.
- Ahmadi, A., Karray, F., & Kamel, M. S. (2010). Flocking based approach for data clustering. *Natural Computing*, 9(3), 767–791.
- Atkinson, G. M., & Boore, D. M. (2003). Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bulletin of the Seismological Society of America*, 93(4), 1703–1729.
- Atkinson, G. M., & Boore, D. M. (2006). Earthquake ground-motion prediction equations for Eastern North America. *Bulletin of the Seismological Society of America*, 96(6), 2181–2205.
- Basu, S., Nigam, N. (1977) "Seismic risk analysis of Indian peninsula." In: *Proceedings of 6th World Conference on Earthquake Engineering*. p 782–790
- Bhatia, S. C., Ravi Kumar, M., & Gupta, H. K. (1999). A probabilistic seismic hazard map of India and adjoining regions. *Annali Di Geofisica*, 42, 1153–1164.
- BIS. (1962). IS 1893–1962: Indian standard recommendations for earthquake resistant design of structures. Bureau of Indian Standards.
- Boore, D. M., Stewart, J. P., Seyhan, E., & Atkinson, G. M. (2014). NGA-west2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra*, 30(3), 1057–1085.
- Campbell, K. W., & Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra*, 30(3), 1087–1115.

- Chiou, B. S. J., & Youngs, R. R. (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, 30(3), 1117–1153.
- Dhanya, J., & Raghukanth, S. T. G. (2020). Neural network-based hybrid ground motion prediction equations for western Himalayas and north-eastern India. *Acta Geophysica*, 68(2), 303–324.
- Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Cybernetics*, 3, 32–57.
- Kennedy, J., & Eberhart, R. (1995) Particle swarm optimization. In *Neural Networks, 1995. Proceeding, IEEE international conference*, 4, (pp. 1942–1948).
- Gardner, J. K., & Knopoff, L. (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bulletin of the Seismological Society of America*, 64(5), 1363–1367.
- Geological Survey of India (GSI) (2000) Seismotectonic Atlas of India and its Environs.
- Goulet, C. A., Bozorgnia, Y., Kuehn, N., Atik, L. A., Youngs, R. R., Graves, R. W., & Atkinson, G. M. (2021). NGA-east ground-motion characterization model part 1: summary of products and model development. *Earthquake Spectra*, 37, 1231–1282.
- Gubin, I. E. (1971). Multi-element seismic zoning (considered on the example of the Indian Peninsula). *Earth Physics*, 12, 10–23.
- Guha, S. K. (1962). Seismic regionalization of India. In *Proceedings of Symposium. Earthquake Engineering*, 2nd. 191–207
- Gupta, I., & Trifunac, M. (2018). Attenuation of strong earthquake ground motion—Dependence on geology along the wave path from the Hindukush subduction to western Himalaya. *Soil Dynamics and Earthquake Engineering*, 114, 127–146.
- IBC. (2015). International building code. *International Code Council, Inc.*
- IS: 1893. (2016). Criteria for earthquake resistant design of structures, Part 1: General provisions and buildings. *Bureau of Indian Standards*, Government of India, New Delhi
- Kale, O. (2019). Some discussions on data-driven testing of ground-motion prediction equations under the Turkish ground-motion database. *Journal of Earthquake Engineering*, 23(1), 160–181.
- Kamalzadeh, H., Ahmadi, A., & Mansour, S. (2017). A shape-based adaptive segmentation of time-series using particle swarm optimization. *Information Systems*, 67, 1–18.
- Kennedy, J., Eberhart, R., & Shi, Y. (2001). *Swarm intelligence*. Morgan Kaufman Publishers.
- Khattri, K. N., Rogers, A. M., Perkins, D. M., & Algermissen, S. T. (1984). A seismic hazard map of India and adjacent areas. *Tectonophysics*, 108(1–2), 93–134.
- Kijko, A., & Graham, G. (1998). Parametric-historic procedure for probabilistic seismic hazard analysis Part I: Estimation of maximum regional magnitude M_{max} . *Pure and Applied Geophysics*, 152(3), 413–442.
- Krishna, J. (1959). Seismic zoning map of India. *Current Science*, 62, 17–23.
- Lapajne, J., Motnikar, B. S., & Zupancic, P. (2003). Probabilistic seismic hazard assessment methodology for distributed seismicity. *Bulletin of the Seismological Society of America*, 93(6), 2502–2515.
- Mohapatra, A. K., and Mohanty, W. K. (2010), “An overview of seismic zonation studies in India”, *Indian Geotechnical Conference*, IGS Mumbai Chapter & IIT Bombay, Dec 16–18, 2010.
- NDMA. (2010). Development of probabilistic seismic hazard map of India. *The National Disaster Management Authority*, Government of India.
- NDMA (2022), “probabilistic seismic hazard map of India”, *The National Disaster Management Authority*, Government of India, New Delhi, https://ndma.gov.in/sites/default/files/PDF/Technical%20Documents/NDMA_PSHMI.pdf. Last accessed on 20/04/2022.
- Parvez, I. A., Vaccari, F., & Panza, G. F. (2003). A deterministic seismic hazard map of India and adjacent areas. *Geophysical Journal International*, 155(2), 489–508.
- Piotrowski, A. P., Napiorkowski, J. J., & Piotrowska, A. E. (2020). Population size in particle swarm optimization. *Swarm and Evolutionary Computation*, 58, 100718.
- Singh, S., Srinagesh, D., Srinivas, D., Arroyo, D., Perez-Campos, X., Chadha, R., Suresh, G., & Suresh, G. (2017). Strong ground motion in the Indo-Gangetic plains during the 2015 Gorkha, Nepal, earthquake sequence and its prediction during future earthquakes. *Bulletin of the Seismological Society of America*, 107(3), 1293–1306.
- Sreejaya, K. P., Raghukanth, S. T. G., Gupta, I. D., Murty, C. V. R., and Srinagesh, D. (2022), “Seismic hazard map of India and neighboring regions”, *Soil Dynamics and Earthquake Engineering* (Under review).
- Tandon, A. N. (1956). Zones of India liable to earthquake damage. *Indian Journal of Meteorology and Geophysics*, 10, 137–146.
- Uhrhammer, R. A. (1986). Characteristics of northern and central California seismicity. *Earthquake Notes*, 57(1), 21.
- Weichert, D. H. (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America*, 70(4), 1337–1346.
- Youngs, R. R., Chiou, S. J., Silva, W. J., & Humphrey, J. R. (1997). Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismological Research Letters*, 68(1), 58–73.
- Yu, Y., Gao, M., & Xu, G. (2011). Seismic Zonation. In H. K. Gupta (Ed.), *Encyclopedia of solid earth geophysics. Encyclopedia of earth sciences series* (pp. 1224–1230). Springer.
- Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., et al. (2006). Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bulletin of the Seismological Society of America*, 96(3), 898–913.

(Received July 8, 2022, revised April 17, 2023, accepted July 12, 2023, Published online August 17, 2023)