Identification of seismically susceptible areas in western Himalaya using pattern recognition

MRIDULA*, AMITA SINVHAL and HANS RAJ WASON

Department of Earthquake Engineering, Indian Institute of Technology Roorkee, Roorkee 247 667, India.

**Corresponding author. e-mail: mridulaitr@gmail.com*

Seismicity in the western Himalayas is highly variable. Several historical and instrumentally recorded devastating earthquakes originated in the western Himalayas which are part of the Alpine–Himalayan belt. Earthquakes cause tremendous loss of life and to the built environment. The amount of loss in terms of life and infrastructure has been rising continuously due to significant increase in population and infrastructure. This study is an attempt to identify seismically susceptible areas in western Himalaya, using pattern recognition technique. An area between latitude 29°–36°N and longitude 73°–80°E was considered for this study. Pattern recognition starts with identification, selection and extraction of features from seismotectonic data. These features are then subjected to discriminant analysis and the study area was classified into three categories, viz., Area A: most susceptible area, Area B: moderately susceptible area, and Area C: least susceptible area. Results show that almost the entire states of Himachal Pradesh and Uttarakhand and a portion of Jammu & Kashmir are classified as Area A, while most of Jammu & Kashmir is classified as Area B and the Indo-Gangetic plains are classified as Area C.

1. Introduction

Western Himalayas are part of the seismically active Alpine–Himalayan belt, and comprises of the states of Jammu & Kashmir (J&K), Himachal Pradesh (HP) and Uttarakhand (UK). The region has experienced several moderate to large-sized earthquakes including the great Kangra earthquake of 1905. In western Himalayas, different areas exhibit different patterns of seismicity, i.e., some parts of Himalaya show concentrated seismicity but some other parts show sparse seismicity. It is, therefore, necessary to differentiate between areas which can experience relatively high seismicity compared to other areas. The areas which experience high seismicity are prone to higher seismic hazards and are defined as seismically susceptible areas. A region bounded by latitude 29°–36°N and longitude 73°–80°E has been considered for this

study and is shown in figure 1(a). An attempt has been made to identify seismically susceptible areas in the western Himalayas using pattern recognition (PR) technique.

Pattern recognition (PR) is the science that concerns either the description and/or classification (recognition) of a set of events. Three important steps identified by Walt and Barnard (2007) are: the study of how machines can observe the environment, how it learns to distinguish patterns of interest from their background, and make sound and reasonable decisions about categories of patterns. The main objective of pattern recognition technique is to recognize objects or patterns in different classes or categories either on the basis of prior knowledge or on the basis of statistical data extracted from patterns. Feature selection and extraction is an important phase of any pattern recognition system. Feature selection is the

Keywords. Western Himalayas; susceptible area; pattern recognition; seismic hazard assessment.

Figure 1. (a) Study area defined by the coordinates latitudes: 29°–36°N and longitudes: 73°–80°E in western Himalaya and (b) seismotectonics of the study area.

process of choosing input to the pattern recognition system and involves judicious judgment. Features may be qualitative, i.e., these can be symbolic (like colour, shape). These may be represented by continuous or discrete-binary variables. On the other hand, features may be represented quantitatively. Features may also be represented by a combination of both qualitative and quantitative features. The environment is observed by a set of features which can be selected or extracted from a given dataset. Related problems of feature selection and extraction are addressed at the beginning of any pattern recognition design. Features which are to be extracted must be able to fulfill the following conditions: extraction of

each feature must be computationally feasible; it must lead to successful recognition of pattern; and it should help in converting the problem data into a manageable amount of information without discarding valuable information (Duda *et al.* 2001). Patterns of interest can be learnt or can be distinguished from their background. This involves evolving and specifying a classification criterion. The classification or description scheme is usually based on availability of a set of patterns that have already been classified or described. To recognize patterns, it is important to train the classifier first by the training dataset that is usually some data that has already been classified or described. Pattern recognition is generally categorized according

to the type of learning procedure used to generate the output value. This then leads to sound and reasonable decisions about categories of patterns.

The earliest known research paper in PR in seismic hazard assessment harks back to 1972. Pattern recognition was used to identify sites in earthquake-prone areas capable of generating events greater than some threshold magnitude by several authors. According to Talwani (1988), the methodology for identifying strong earthquakes is based on the assumption that large earthquakes are related to intersections of lineaments, which provide a location for stress accumulation. Talwani (1999), proposed a model demonstrating that intersecting faults provide a location for stress accumulation. Also, according to King (1986), fault intersections provide locations for initiation and cessation of ruptures. An algorithm, CORA-3, was developed for identification of earthquake-prone areas by Gelfand *et al.* (1972) and was modified later and used by several authors for identification of earthquake-prone areas for different regions in the world. Further developments of pattern recognition techniques in seismic hazard assessment are shown in table 1. Methodologies given in table 1 rely on an analysis of available geological and morphostructural data, founded on a set of parameters

based on relief, topography, drainage, geology, sediments and faults. These methods used hierarchical logical decision making, which is essentially a qualitative pattern recognition technique, for morphostructural zoning and then to recognize seismogenic nodes. Therefore, these methodologies are independent of past seismicity and are particularly useful in areas where information on historical seismicity is sparse (Peresan *et al.* 2015). However, seismogenic nodes that are not related with known active faults or past earthquakes still need investigation. In addition, epicentral data, though used in analysis, was not a discriminant parameter. This feature was taken care of and enhanced by the PR technique developed by Sinval *et al.* (1990, 1991), as a quantitative approach was applied and epicentral data was part of extracted features in the PR analysis. Three distinct types of zones for Tehri region were identified by PR method which was based on linear discriminant analysis. The features identified were based on tectonic density, drainage density and seismic density. These were:

- magnitude of earthquake;
- number of major thrusts;
- distance to end of major thrust;
- number of lineaments (identified through satellite imageries);

Table 1. *PR studies for identification of earthquake prone areas (after Mridula et al. 2013).*

Author(s)	Year	Area	Result
Gelfand <i>et al.</i>	1972	Eastern central Asia	Six knots were identified where strong shocks may occur
Gelfand <i>et al.</i>	1974	Anatolia adjacent	Not defined
Gelfand <i>et al.</i>	1976	California	The dangerous areas identified were characterized by proximity intersections or ends of major faults and by relatively low elevation or subsidence against a background of weaker uplift
Varunoday <i>et al.</i>	1979	Kumaon Himalayan	Out of the 51 objects considered, 33 appear as dangerous and 18 as non-dangerous
Caputo <i>et al.</i>	1980	Italy	Two clusters of dangerous points were observed in north and central Apennines
Bhatia <i>et al.</i>	1992	Himalayan Arc	Two studies were carried out for $M \geq 6.5$ and $M \geq 7.0$. Out of 97 knots 48 knots are seismically potential for $M \geq 6.5$ and 36 knots are seismically potential for $M \geq 7.0$.
Gorshkov <i>et al.</i>	2009	Alborz region	Of 134 nodes, 79 (58%) are classified as dangerous and 65 (49%) as non-dangerous
Gorshkov <i>et al.</i>	2010	Iberian peninsula	Most of the dangerous nodes are scattered at the periphery of the peninsula
Peresan <i>et al.</i>	2011	Northeastern Italy	Dangerous nodes were identified as the circles of radius 25 km

- number of intersections of lineaments;
- length of river course/tributary;
- number of epicenters.

Three distinct types of zones were identified: S1 was identified as a highly critical zone, approximately 100 km² in area, with three major thrusts within the area: Krol, Garhwal and Tons thrusts. Narenda Nagar, Devprayag, Kirtinagar, Chamba, Jhaknidhar and Rishikesh are within this critical zone. S2 was identified as a moderately critical zone and existed in three separate islands, and was marked by two major thrusts: Bhatwari and Munsiari thrust. The three islands were centred around: (a) Badhani, Bhudna; (b) between North Almora thrust and Tons Nayar thrust, centred at Tehri; and (c) Garhwal thrust and Krol thrust, centred at Rishikesh. S3 was identified as a critical zone, and existed along all the major thrusts.

A hypothetical disaster scenario together with vulnerability and risk analysis was conceived for Narendranagar in the highly critical zone, S1, by Gupta *et al.* (2006). Fifty-nine percent of the population of Narendranagar was found to be vulnerable to damage associated with higher accelerations of 0.41 g. This risk increased if the earthquake magnitude was larger, and could be even higher in the vicinity of faults, intersection of fault and river and in areas of higher population density. Forty-seven villages and one urban center with 32.4% of the total population of Narendranagar were at higher risk due to seismotectonics of the region.

Hypocenter of the modelled Uttarkashi earthquake was in the moderately critical zone, S2, and that of the Chamoli earthquake of 1999 was below an extension of the same zone; implying that S2 seems to have tremendous seismogenic potential (Gupta and Sinval 2010; Sinval 2010).

With this background, therefore, the objective of the present study is to divide the western Himalayas into several seismically susceptible areas and then to validate these areas according to their seismogenic potential. For this a quantitative approach was applied, using a PR technique. Features were identified and extracted from the available tectonic and seismicity data of the study area. These features were in terms of epicentral density, tectonic density, seismotectonic density and elevation data. A classification criterion was formulated and tested. Seismicity data was compiled from various earthquake catalogues, tectonic data was studied and compiled from 'Seismotectonic Atlas of India and its Environ' (Narula *et al.* 2000), elevation data was taken from the digital elevation model provided by Indian Space Research Organization (ISRO). Details of these data are given in the following sections.

1.1 Seismicity of the region

Seismicity of the region is attributed mainly to the convergence between the Indian and the Eurasian plates, with continent–continent collision. The region has a history of experiencing moderate to large magnitude destructive earthquakes, most with shallow depths, less than 20 km. To consider seismicity of the region, a comprehensive seismicity catalogue is essential, which contains hypocentral parameters, time of occurrence, and magnitude. This data is required for as long a time period as possible, i.e., from the earliest times to the present. This data is not available in a single catalogue. Therefore, various earthquake catalogues from different sources were searched, accessed, selected, collated and subsequently studied for several aspects like chronological continuity, magnitude types and main shocks, foreshocks and aftershocks. These catalogues included India Meteorological Department (IMD), International Seismological Center (ISC), and United States Geological Survey (USGS) data, which were merged. The purpose of merging is to remove duplicate earthquake events from the merged catalogue. The IMD catalogue was considered for historical seismicity, starting from the year 1552 and continuing up to 1963, magnitude ranged from $M_L = 4.0$ to 8.0 and 89 earthquakes emerged for the study area. The ISC catalogue was considered for the years between 1964 and 2010, where the magnitude ranged from $m_b = 3.0$ to 6.7 for all events, and it had the largest number of events, 2619. Beyond 2010, the USGS (National Earthquake Information Center, NEIC) catalogue was considered, in which 41 recent earthquakes were considered. The magnitude ranged from $m_b = 4.0$ to 5.1.

The merged catalogue consisted of 2749 events within the study area. The lower limit of magnitude in this catalogue was 3.0, and it was on different scales, such as m_b and M_L . For the present analysis, it was necessary that all magnitudes are homogenized to a single scale. Therefore, the merged catalogue was homogenized and all magnitudes were converted to moment magnitude, M_w . M_L was considered equivalent to moment magnitude as per Ristau *et al.* (2005) and m_b was converted to M_w by using general orthogonal relations given by Wason *et al.* (2012). Henceforth, moment magnitude, M_w , is referred as magnitude in this paper.

After merging and homogenization, the earthquake catalogue was subjected to declustering, using the Gardner and Knopoff (1974) method. After subjecting the compiled events to the process of merging, homogenization and declustering, a new workable catalogue emerged with 1172 earthquakes, magnitude ranged from $M_w = 3.5$ to 8.0, for the period 1552–2012. The Merged,

Homogenized and De-clustered catalogue is henceforth referred as the MHD earthquake catalogue for Western Himalayas.

1.2 Completeness of earthquake data

Analysis of available seismicity dataset to assess the nature and degree of its completeness is an important step for any seismicity analysis. Therefore, in seismicity study where statistical assumptions have a major role, it becomes essential to assess the completeness of the dataset. Even though historical seismicity is sparse, and most of the time, it is quite inaccurate in terms of magnitude and epicentral location, yet it is necessarily used in seismic hazard assessment. Instrumental records of earthquakes are often too short in time, and reliable ones exist only after 1964. Most catalogs are inconsistent with respect to time and space. Stepp's (1972) method has been used in this study to analyze the time completeness of the MHD catalogue. In order to analyze the nature of the completeness of earthquake dataset, the dataset was first grouped into different magnitude classes and then each magnitude class was modelled as a point process in time. In this method, the sample mean is inversely proportional to the number of observations in the sample. Rate of occurrence as a function of time interval for different magnitude classes was calculated. To obtain an estimate of the sample mean it was assumed that the earthquake sequence can be modelled as a Poisson distribution. Table 2 shows the rate of

occurrence of earthquakes of different magnitudes as a function of time. Rate is given as N/T , where N is the cumulative number of earthquakes in the time interval T . Figure 2 shows the time distribution of the yearly number of earthquakes for different magnitude ranges as per table 2. This data is then used to compute the standard deviation of the estimates of mean, as given in table 3, using equation: $\sigma_\lambda = (\lambda/T)^{1/2}$, where T is time interval, λ is the mean rate per unit time interval and is given as: $\lambda = (1/n) \sum_{i=1}^n k_i$, $k_1, k_2, \dots, k_n$ are the number of earthquakes per unit time interval. The standard deviation of the mean rate as a function of sample period is plotted along with nearly tangent lines with slopes $1/\sqrt{T}$, and is shown in figure 3. This indicated the period up to which a particular magnitude range may be taken to be complete. The completeness period for different magnitude ranges were estimated as follows: ≤ 3.5 for 32 years, 3.6–4.0 for 52 years, 4.1–4.5 for 52 years, 4.6 to 5.0 for 60 years, 5.1 to 5.5 for 92 years, 5.6–6.0 for 112 years and ≥ 6.1 for 232 years.

Magnitude of completeness (M_c) for the MHD catalogue was computed using maximum curvature method (Wyss *et al.* 1999; Wiemer and Wyss 2000). For calculating magnitude of completeness of the MHD catalogue cumulative number of earthquakes *versus* magnitude was plotted. Figure 4 shows the frequency magnitude distribution (FMD) of earthquakes of MHD catalogue for the study area. The magnitude of completeness for the entire MHD catalogue was estimated as $M_w = 4.3$. However, magnitude in the MHD catalogue

Table 2. Rate of earthquake occurrence of different magnitude ranges and time intervals.

Time period	Time interval (years)	Magnitude (M_w)													
		≤ 3.5		3.6–4.0		4.1–4.5		4.6–5.0		5.1–5.5		5.6–6.0		≥ 6.1	
		N	N/T	N	N/T	N	N/T	N	N/T	N	N/T	N	N/T	N	N/T
2001–2012	12	5	0.42	85	7.08	290	24.17	74	6.17	15	1.25	6	0.50	1	0.08
1991–2012	22	9	0.41	124	5.64	441	20.05	165	7.50	23	1.05	8	0.36	4	0.18
1981–2012	32	12	0.38	125	3.91	472	14.75	275	8.59	34	1.06	12	0.38	5	0.16
1971–2012	42	14	0.33	127	3.02	486	11.57	346	8.24	63	1.50	16	0.38	8	0.19
1961–2012	52	15	0.29	128	2.46	489	9.40	371	7.13	82	1.58	18	0.35	8	0.15
1951–2012	62	15	0.24	128	2.06	489	7.89	371	5.98	84	1.35	20	0.32	10	0.16
1941–2012	72	15	0.21	128	1.78	489	6.79	371	5.15	86	1.19	23	0.32	13	0.18
1931–2012	82	15	0.18	128	1.56	489	5.96	372	4.54	88	1.07	24	0.29	14	0.17
1921–2012	92	15	0.16	128	1.39	489	5.32	372	4.04	92	1.00	24	0.26	15	0.16
1911–2012	102	15	0.15	128	1.25	489	4.79	372	3.65	92	0.90	25	0.25	15	0.15
1901–2012	112	15	0.13	128	1.14	489	4.37	373	3.33	92	0.82	27	0.24	18	0.16
1891–2012	122	15	0.12	128	1.05	489	4.01	373	3.06	92	0.75	27	0.22	18	0.15
1881–2012	132	15	0.11	128	0.97	489	3.70	373	2.83	92	0.70	28	0.21	21	0.16
1871–2012	142	15	0.11	128	0.90	489	3.44	374	2.63	93	0.65	29	0.20	21	0.15
1811–2012	202	15	0.07	128	0.63	489	2.42	378	1.87	95	0.47	31	0.15	24	0.12
1711–2012	302	15	0.05	128	0.42	489	1.62	378	1.25	96	0.32	32	0.11	29	0.10
1611–2012	402	15	0.04	128	0.32	489	1.22	378	0.94	96	0.24	32	0.08	32	0.08
1552–2012	461	15	0.03	128	0.28	489	1.06	378	0.82	96	0.21	32	0.07	34	0.07

Figure 2. Time distribution of the yearly number of earthquakes for different magnitude ranges.

Table 3. Standard deviation (σ_λ) of the estimates of mean recurrence rate per year (λ).

Time interval (years)	Magnitude (M_w)													
	< 3.5		3.6–4.0		4.1–4.5		4.6–5.0		5.1–5.5		5.6–6.0		> 6.1	
	N/T	σ	N/T	σ	N/T	σ	N/T	σ	N/T	σ	N/T	σ	N/T	σ
12	0.42	0.19	7.08	0.77	24.17	1.42	6.17	0.72	1.25	0.32	0.50	0.20	0.08	0.08
22	0.41	0.14	5.64	0.51	20.05	0.95	7.50	0.58	1.05	0.22	0.36	0.13	0.18	0.09
32	0.38	0.11	3.91	0.35	14.75	0.68	8.59	0.52	1.06	0.18	0.38	0.11	0.16	0.07
42	0.33	0.09	3.02	0.27	11.57	0.52	8.24	0.44	1.50	0.19	0.38	0.10	0.19	0.07
52	0.29	0.07	2.46	0.22	9.40	0.43	7.13	0.37	1.58	0.17	0.35	0.08	0.15	0.05
62	0.24	0.06	2.06	0.18	7.89	0.36	5.98	0.31	1.35	0.15	0.32	0.07	0.16	0.05
72	0.21	0.05	1.78	0.16	6.79	0.31	5.15	0.27	1.19	0.13	0.32	0.07	0.18	0.05
82	0.18	0.05	1.56	0.14	5.96	0.27	4.54	0.24	1.07	0.11	0.29	0.06	0.17	0.05
92	0.16	0.04	1.39	0.12	5.32	0.24	4.04	0.21	1.00	0.10	0.26	0.05	0.16	0.04
102	0.15	0.04	1.25	0.11	4.79	0.22	3.65	0.19	0.90	0.09	0.25	0.05	0.15	0.04
112	0.13	0.03	1.14	0.10	4.37	0.20	3.33	0.17	0.82	0.09	0.24	0.05	0.16	0.04
122	0.12	0.03	1.05	0.09	4.01	0.18	3.06	0.16	0.75	0.08	0.22	0.04	0.15	0.03
132	0.11	0.03	0.97	0.09	3.70	0.17	2.83	0.15	0.70	0.07	0.21	0.04	0.16	0.03
142	0.11	0.03	0.90	0.08	3.44	0.16	2.63	0.14	0.65	0.07	0.20	0.04	0.15	0.03
202	0.07	0.02	0.63	0.06	2.42	0.11	1.87	0.10	0.47	0.05	0.15	0.03	0.12	0.02
302	0.05	0.01	0.42	0.04	1.62	0.07	1.25	0.06	0.32	0.03	0.11	0.02	0.10	0.02
402	0.04	0.01	0.32	0.03	1.22	0.06	0.94	0.05	0.24	0.02	0.08	0.01	0.08	0.01
461	0.03	0.01	0.28	0.02	1.06	0.05	0.82	0.04	0.21	0.02	0.07	0.01	0.07	0.01

ranged between 3.5 and 8.0, and magnitudes lower than the magnitude of completeness, $M_w = 4.3$, were retained in this study as these were considered as discriminant features in subsequent analysis.

1.3 Tectonics of the region

Due to continent–continent collision of the Indian plate with the Eurasian plate, the study area is manifested with complex tectonics. For this study tectonics of the area was considered as per ‘Seismotectonic Atlas of India and its Environ’ (Narula *et al.* 2000). Several thrusts and faults are parallel and subparallel to the Himalayan arc. From north to south, these are: the Indus Suture Zone (ISZ), Main Central Thrust (MCT), Main

Boundary Thrust (MBT), and the Main Frontal Thrust (MFT). These mega tectonic units manifest throughout the Himalayan arc, and have prominent surface manifestations at several places in western Himalayas. The MCT terminates against the Kishtwar Fault in Jammu and Kashmir. MFT is also known as the Frontal Foothill Thrust (FFT) and is a neo-tectonic thrust. Besides these mega Himalayan thrusts, several other faults and thrusts exist in the study area. The Karakoram Fault (KF), north of the ISZ, exhibits a huge offset and extends for more than 1000 km from central Pamir to north of Uttarakhand Himalayas. Seven complex tectonic units exist in the Lesser Himalaya, in the form of closed looped thrusts, between the MBT and the MCT. The region between the MBT and

Figure 3. Variation of σ_λ with time interval and magnitude and lines with slope $(1/\sqrt{T})$ for testing completeness of earthquake data of the study area.

Figure 4. Frequency magnitude distribution (FMD) for estimation of minimum magnitude of completeness of homogenized catalogue.

the FFT is traversed by several thrusts and faults. Some of these, such as the Jwalamukhi Thrust (JT) and Drang Thrust (DT), can be traced over long distances. In addition to the tectonic features mentioned here, several prominent faults and lineaments are transverse to the Himalayan trend, such as the Kishtwar Fault, Sundarnagar Fault (SNF), and the Mahendragarh–Dehradun (MHD–DDN F) subsurface fault. In all, 118 tectonic elements were identified (Narula *et al.* 2000) and digitized in ArcGIS-9.3.

The seismotectonic map, shown in figure 1(b), emerged by overlaying the merged, homogenized and declustered MHD earthquake catalogue data on to the tectonic map of the region. Seismotectonics of the region showed a very interesting pattern. Seven clusters of dense seismicity were observed. These are: around the western syntaxes (epicenter of the Kashmir earthquake of 8th October, 2005;

$M_w = 7.6$ (MHD), was part of this cluster); along the Great Karakoram Range; in the area bound by the MBT in the south; MCT in the north; Kishtwar Fault in the west; Sundarnagar Fault in the east; epicenter of the great Kangra earthquake of 1905 is a part of this dense cluster and it is in the vicinity of the MBT; and a prominent cluster of seismicity lies east of the study area, along the MCT. Uttarkashi earthquake of 20th October 1991, $M_w = 6.8$ (MHD) and Chamoli earthquake of 28th March 1999, $M_w = 6.7$ (MHD), are part of this cluster. Another prominent cluster was observed transverse to the Himalayan trend, along the Kaurik Fault system. The Kinnaur earthquake of 19th January 1975, $M_w = 6.6$ (MHD) is part of this cluster. Another cluster of dense seismicity was observed east of Karakoram Fault and along Beng Co Fault in the northeastern part of the study area. Compared to the regions already discussed, seismicity was found to be sparse in three regions, viz., south of the MBT, along the Jhelum Fault, and in the region along ISZ, roughly defined between MCT, Shyok Suture, Kishtwar Fault and Tso Morari Fault.

Mukhopadhyay *et al.* (2011) observed five spatial clusters in the vicinity of MCT in the entire Himalayan arc, and put forward the hypothesis that MCT is seismically active in segments. This aspect was used for segmentation of MCT (Mridula *et al.* 2016, communicated). Seismogeological sections along Nanga Parbat syntaxes in the western Himalaya show a depth penetration of more than 200 km of the Indian plate (Mukhopadhyay 2011). Dasgupta *et al.* (2013) illustrated that transverse tectonic features play an important role in the Himalayan collision, by studying the Sikkim earthquake of 18th September, 2011 and Bihar–Nepal earthquake of 20th August, 1988.

1.4 Elevation data

Digital elevation model (DEM) available on National Remote Sensing Centre (Indian Space Research Organization, ISRO) website was considered for the study area. ISRO launched Cartosat-1 with the primary objective of providing high resolution satellite data of 2 m in track stereo. DEM generated by Cartosat-1 facilitates large scale mapping and terrain modelling. A digital elevation model (DEM) is a digital model or 3-dimensional representation of a terrain's surface. DEM can be represented as a raster (a grid of squares, also known as a height map when representing elevation) or as a vector-based triangular irregular network (TIN). DEMs were downloaded from the website (<http://bhuvan.nrsc.gov.in/bhuvan.links.php>) and elevation contours were generated in ArcGIS-9.3.

2. Methodology

Methodology consists of two parts. In the first part, epicentral data of the MHD catalogue was classified into two classes on the basis of magnitude. An area of influence of 25 km was drawn around each epicenter. A set of features, which were identified for each earthquake, were extracted from within this circle. Epicenters of both classes were then subjected to discriminant analysis. The objective of discriminant analysis is to find a linear combination of several features that produces the maximum difference between the two previously defined classes. If we can find a function that produces a significant difference, we can attempt to allocate any new specimen of unknown origin to one of the two original classes (Davis 2002). Several classification criteria and several sets of features were tested for PR and the set which gave the least error in terms of misclassification of epicenters was retained for further study. Coefficient of discriminant function for each feature and seismic score for each epicenter was computed. This consists of the training part.

In the second part, i.e., in decision making, the discriminant function as obtained in the training part was used to classify the unknown into either of the two classes, i.e., the next step is to identify susceptible areas in the study area. The study area was divided into a grid of 0.5×0.5 degree. An area of influence of 25 km was drawn around each grid point, as in part 1. The same set of features which were identified earlier were extracted and seismic score was calculated for each grid point. On the basis of seismic score, these grid points were classified into different susceptible areas which were later named as class A, B or C. Details of both

steps of methodology are explained in the following sections.

2.1 Classification criteria

All 537 epicenters in the MHD catalogue, $4.5 \leq M_w \leq 8.0$, were classified on the basis of magnitude as training data. Earthquakes below magnitude $M_w = 4.5$ were not considered as the magnitude of completeness of the entire MHD catalogue was 4.3, and moreover magnitudes below this range are not of much engineering significance. In an initial attempt at formulating a classification criteria epicenters with magnitude $M_w \geq 6.0$ were grouped as class A and consisted of 45 events, and epicenters with magnitude $4.5 \leq M_w \leq 5.9$ were grouped as class B and consisted of 492 events.

2.2 Feature selection and extraction

A circle of radius 25 km, centered on the epicenter was drawn around each of the 537 epicenters. The selected epicenter is referred to as the central earthquake henceforth. All features were identified and quantified within this radius, using software package ArcGIS 9.3. Using seismicity data from the MHD catalogue and 118 tectonic units, eleven features which included the seven features identified by Sinvhal *et al.* (1990, 1991), were identified and extracted from the circle of radius 25 km around each epicenter. The features used in this study are: F1: number of epicenters, F2: number of tectonic units, F3: number of tectonic intersections, F4: distance between central earthquake and nearest tectonic unit (km) F5: distance between central earthquake and nearest tectonic intersection (km), F6: length of river (km), F7: number of intersections of tectonic units with river, F8: number of earthquakes + number of tectonic units (F1 + F2), F9: minimum elevation ($E_{\min}$), F10: maximum elevation ($E_{\max}$), and F11: difference in elevation ($E_{\max} - E_{\min}$).

Features F1–F8 were extracted from seismicity and tectonic data whereas F9–F11 were extracted from elevation data. These features are shown for the great Kangra earthquake of 1905 ($M_w = 8.0$), in figure 5(a and b). It shows that there are eight epicenters within the circle besides the epicenter of Kangra earthquake. Six tectonic units are MBT, Mastgarh Anticline, Jwalamukhi Thrust, Drang Thrust, unnamed thrust TH 07 and lineament L3. The lineament L3 intersects MBT, TH 07 and JMT and Drang Thrust intersects MBT, therefore there are four tectonic intersections. The central earthquake is 0.14 km from the nearest tectonic feature

TH 07 and is at 15.75 km from the nearest tectonic intersection. The Ravi River has a length of 25.68 km within the circle. There is no intersection between the river and tectonic units. Feature F8 is the combination of tectonics and earthquake epicenters within the circle which is equal to 14. Maximum elevation is 4850 m and minimum elevation is 350 m, therefore difference in elevation is 4500 m. Similarly, all 11 features were extracted for all 537 epicenters in the MHD catalogue in the range $4.5 \leq M_w \leq 8.0$. These 11 features were used in all subsequent exercises in discriminant analysis.

2.3 Discriminant analysis

With this classification criterion and with 11 features mentioned above, the first exercise in discriminant analysis was carried out. These 537 epicenters were then subjected to discriminant analysis. A linear combination of extracted features was used to maximize the difference between multivariate means of linear combination of all 11 features of class A and B. A discriminant function was then developed that transformed the original set of features of an earthquake into a single

Figure 5. (a) An example of extracted features for the Great Kangra earthquake of 1905, $M_w = 8.0$ (MHD catalogue); (b) Elevation contours as per ISRO website within the circle around the Great Kangra earthquake of 1905, at 500 m interval. Minimum elevation is shown by green and maximum elevation is shown by red triangle. Table shows the numerical values of the extracted features. F1, F2, ..., F11 are the extracted features and are described in section 2.2.

seismic score. Coefficients of discriminant function, $\lambda_1, \lambda_2, \dots, \lambda_{11}$, were calculated for each feature and seismic score was calculated for each earthquake, which is the sum of products of λ of each feature to the numerical value of that feature, i.e., $\lambda_1 * F1 + \lambda_2 * F2 + \dots + \lambda_{11} * F11$. Multivariate mean of class A (R_A) and of class B (R_B) and discriminant index (R_0), which is exactly the halfway between the R_A and R_B , i.e., the mean of R_A and R_B were computed.

When plotted on a graph, these values were used to check whether each epicenter was properly classified or not, i.e., whether seismic score of any earthquake of class A was located on the class A side of R_0 or not, and similarly this was checked for class B. In the subsequent graphical exercise out of 45 seismic scores of class A, 43 were on class A side and two were on class B side, i.e., 95% epicenters of higher magnitude class were desirably classified. However, for class B out of 492

earthquakes 324 were on class B side and 168 epicenters were on class A side, i.e., class B showed 34% misclassification.

To reduce this misclassification, a second classification criterion was formulated and tested in which to increase the number of events in class A, a lower magnitude threshold was considered. 537 epicenters of magnitude $4.5 \leq M_w \leq 8.0$ from MHD catalogue were re-classified into two training classes, class A and class B. Class A consisted 94 epicenters which had magnitude $M_w \geq 5.5$, and class B consisted 443 epicenters which had magnitude in the range $4.5 \leq M_w \leq 5.4$. With this classification criterion and with different combinations of 11 features, 15 iterations of discriminant analyses were carried out, and percentage of misclassification was computed for all iterations, as given in table 4. It was observed that a complete set of 11 features gave the most desirable result in terms of percentage of misclassification, given at

Table 4. *Percentage of classification of either class by varying classification criteria and number of features.*

Sl. no.	Classification of magnitude		No. of epicenters in class A	No. of epicenters in class B	Number of features used	% Misclassification of class A	% Misclassification of class B
	Class A	Class B					
1.	$M_w \geq 6.0$	$4.5 \leq M_w \leq 5.9$	45	492	11 (F1–F11)	5	34
2.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F1, F2, F3, F4	8	15
3.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	7 features (F1, F2, F3, F4, F6, F9, F10)	4	18
4.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	11 (F1–F11)	0	9
5.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F1 (Number of earthquakes) excluded	8	22
6.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F2 (Number of tectonic units) excluded	6	19
7.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F3 (Number of tectonic intersections) excluded	3	7
8.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F4 (Distance of central earthquake from nearest tectonic feature) excluded	2	12
9.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F5 (Distance of central earthquake from nearest tectonic intersection) excluded	3	12
10.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F6 (Length of river) excluded	1	14
11.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F7 (Number of tectonic river intersections) excluded	0	11
12.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F8 (Combination of no. of earthquake and no. of tectonic features) excluded	4	15
13.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F9 (Minimum elevation) excluded	4	8
14.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F10 (Maximum elevation) excluded	3	8
15.	$M_w \geq 5.5$	$4.5 \leq M_w \leq 5.4$	94	443	F11 (Difference in elevation (Max. elevation – min. elevation)) excluded	0	10

sl. no. 4 in table 4 and was retained for identifying susceptible areas. Coefficient of discriminant function for each feature was considered as the training part and the second part was used to classify the unknown into either of the two classes, i.e., which was subsequently used to identify susceptible areas. Coefficients of discriminant function, $\lambda_1, \lambda_2, \dots, \lambda_{11}$, calculated for each feature, are given in table 5. Figure 6 shows discrimination of seismic scores, and the values of R_A , R_B and R_0 as -0.551 , -128.14 and -64.34 , respectively. Seismic scores of all epicenters of class A were on class A side of R_0 which shows a desirable classification, i.e., 100%. However, for class B, 40 seismic scores out of 443 were misclassified, i.e., class B showed 9% misclassification.

The λ values obtained from the training exercise were subsequently used to assign a class to each grid point, whether A or B. To identify seismically susceptible areas, the study area was divided into a grid of 0.5×0.5 degree which gave 196 grid points. The midpoint of each grid rectangle was taken as

Table 5. Coefficient of discriminant function, λ , calculated for features F1 to F11.

Features	λ	Value
Number of earthquake epicenters	λ_1	-3.677
Number of tectonic units	λ_2	-3.699
Number of tectonic intersections	λ_3	1.1754
Distance of central earthquake from nearest tectonic feature	λ_4	0.0769
Distance of central earthquake from nearest tectonic intersection	λ_5	-0.051
Length of river	λ_6	0.0336
Number of tectonic river intersections	λ_7	-0.104
Combination of no. of earthquake and no. of tectonic features	λ_8	3.561
Minimum elevation	λ_9	-4.548
Maximum elevation	λ_{10}	0.264
Difference in elevation (Max. elevation - min. elevation)	λ_{11}	6.8140

the centre of a circle, of radius of 25 km, as shown in figure 7 and the same eleven features, F1–F11, were extracted for each grid point.

Using the λ values obtained in the training set seismic score computed for each grid point was compared with the R_A , R_B and R_0 values obtained in the training exercise. Seismic scores of grid points falling on R_A side of R_0 were assigned to class A and those on R_B side of R_0 were assigned to class B. At certain grid points very few of the 11 features existed, and the seismic score belonged to neither class A nor B. To accommodate these, therefore, a new class C was introduced. The study area is now divided into three classes, class A, class B and class C.

Figure 7. 196 grid points; each circle represents an area of 25 km radius drawn around each grid point for extraction of features. Colour shows class A, class B and class C after classification.

Figure 6. Seismic scores for earthquake epicenters, R_A , R_B and R_0 for training samples. Blue represents seismic scores for class A and red represents class B.

Out of 196 grid points, 56 grid points were classified as class A; 71 as class B and 69 as class C. These are shown in figure 7. Envelopes were drawn around clusters of grids of same class, using nearest neighbour interpolation tool given in ArcGIS-9.3, which approximates the value of a function for a non-given point when given the value of that function in points around (neighbouring) that point. The nearest neighbour algorithm selects the value of the nearest point and does not consider the values of neighbouring points at all, yielding a piecewise-constant interpolant. Three areas were identified from these clusters, and these were named according to their class as A, B or C. Figure 8 shows the three identified areas. Red represents Area A, yellow represents Area B and green represents Area C. Area A represents the area having $M_w \geq 5.5$; therefore, it may be considered as the most susceptible area. Seven islands of Area A exist in the study area. Area B surrounds Area A, and represents the area having $4.5 \leq M_w \leq 5.4$, therefore, it may be considered as an area less susceptible than Area A.

3. Results and interpretation

To validate these results recent earthquake data was considered, which was not part of the MHD catalogue, for the entire study area and for magnitude ranges considered in this study. The USGS catalogue from 2013 to 2015 showed 70 earthquakes

in the magnitude range $4.0 \leq M_w \leq 5.7$ in the study area. Out of these 70 earthquakes four were of magnitude $M_w \geq 5.5$, 26 were in magnitude range $4.6 \leq M_w \leq 5.4$ and 40 earthquakes were in magnitude range $4.5 \leq M_w \leq 4.0$. It was observed that out of these 70 earthquakes, 44 were located within Area A, i.e., most of the recent earthquakes are occurring in Area A. Moreover, it is significant to note that all four earthquakes of higher magnitude, i.e., $M_w \geq 5.5$ are also located within Area A. It is pertinent to note that this fulfills not only the rigors of classification criteria but also set of features employed. Therefore, success rate for validation for Area A may be considered as 100% in terms of magnitude. Therefore, Area A is validated as the most susceptible area within the study

Table 6. Magnitude wise break up of 70 earthquakes from USGS taken for validation of results.

Magnitude range↓	Area ↓			Total
	A	B	C	
≥ 5.5	4	0	0	4
4.6–5.4	18	7	1	26
≤ 4.5	22	12	2	40
	Σ 44	Σ 19	Σ 3	Σ 70

Figure 8. Three susceptible areas identified after classification. Red represents class A: most susceptible area; Yellow represents class B: moderately susceptible area; and green represents class C: least susceptible area. Places as given in SEISAT are also shown.

Figure 9. Recent earthquakes from USGS and PESMOS plotted on areas for validation. Earthquakes in purple are from PESMOS and blue are from USGS.

Table 7. States and districts in identified susceptible Areas A, B and C.

Area	States	Districts
A	Jammu & Kashmir	Anantnag, Doda, Jammu, Kishtwar, Kulgam, and Reasi
	Himachal Pradesh	Bilaspur, Chamba, Kangra, Kullu, Lahaul & Spiti, Mandi, Shimla, Solan and Sirmaur
	Uttarakhand	Almora, Chamoli, Dehradun, Haridwar, Nainital, Pauri Garhwal, Tehri Garhwal and Uttarkashi
	Punjab	Mohali
	Uttar Pradesh	Bijnor
	Haryana	Panipat and Sirsa
B	Jammu & Kashmir	Bandipora, Kargil, Kathua, Leh (Ladakh), Srinagar, and Udhampur
	Himachal Pradesh	Parts of Bilaspur, Chamba, Kinnaur, Kangra, Lahaul & Spiti, Shimla, Sirmaur, and Una
	Uttarakhand	Almora, Bageshwar, Dehradun, Haridwar, and Uttarkashi
	Punjab	Bathinda, Faridkot, Gurdaspur and Ludhiana
	Uttar Pradesh	Muzaffarnagar
	Haryana	Ambala, Fatehabad, Hisar, Jind, Kaithal, Karnal and Yamunanagar
C	Rajasthan	Ganganagar and Hanumangarh
	Jammu & Kashmir	Kargil and Leh (Ladakh)
	Himachal Pradesh	Una
	Uttar Pradesh	Bijnor, Muzaffarnagar and Sharanpur
	Punjab	Amritsar, Firozpur, Gurdaspur, Hoshiarpur, Kapurthala, Ludhiana, Mansa, Moga, Patiala, Sangrur and Tarn Taran
	Rajasthan	Ganganagar and Hanumangarh

Figure 10. Seismic zoning map by BIS superimposed on identified susceptible areas. Red represents the seismic zone V of BIS.

area. This also implies that recent higher magnitude earthquakes are originating in identified susceptible Area A, and it can be considered as a source zone where future large earthquakes can be expected. Occurrence of smaller magnitude earthquakes together with higher magnitudes in Area A implies that Area A is seismically most active

Figure 11. Identified susceptible areas superimposed on seismic source zones identified in PSHA study (Mridula *et al.* 2014).

compared to other areas. This implies the need of special attention for hazard assessment in Area A. A similar exercise was carried out for validation of Area B and 19 out of 70 earthquakes were

located in Area B. It may be noted that no earthquake of magnitude $M_w \geq 5.5$ occurred in Area B. Similarly three earthquakes of lower magnitude were located in Area C. Similarly, all eight earthquakes from PESMOS (Programme for Excellence in Strong Motion Studies) data from 2013 to 2014 in the magnitude range $4.0 \leq M_w \leq 5.2$ were in Area A. Magnitude-wise distribution of these earthquakes is given in table 6 and is shown in figure 9.

The study area is, therefore, divided into three seismically susceptible areas, where Area A is the most susceptible area, Area B is a moderately susceptible and Area C is the least susceptible area. Large parts of Himachal Pradesh and Uttarakhand and parts of Jammu and Kashmir are the most susceptible areas. In Himachal Pradesh, major parts of Bilaspur, Chamba, Kangra, Kullu, Lahaul & Spiti, Mandi, Shimla, Solan and Sirmaur are within Area A. In Uttarakhand, parts of Almora, Chamoli, Dehradun, Haridwar, Nainital, Pauri Garhwal, Tehri Garhwal and Uttarkashi are in Area A. Districtwise breakup of the three kinds of susceptible areas within the western Himalaya and the Indo-Gangetic plains is given in table 7.

The susceptible area map when superimposed on the seismic zoning map of India (IS: 1893 (Part 1)-2002) and compared with it reveals the following, as shown in figure 10. It was observed that epicenters of recent earthquakes (USGS and PESMOS) lie mostly outside and between the three islands of zone V in western Himalaya, i.e., these are located mostly in seismic zone IV of BIS map. Area A (susceptible area map) is continuous along the entire MCT, and extends north of MCT, sometimes as much as 50 km, and a prominent 50 km by 100 km finger cuts out along the Kaurik Fault system. Large portions of MBT are also included in Area A, which extends south of MBT also, and it also includes areas between MBT and MCT. It implies that zone V in BIS map needs modification in western Himalaya and it is recommended that it should be considered as a continuous and larger area between and along the MBT and MCT, extending from J&K, HP and Uttarakhand as shown by Area A in figure 10. On similar lines as western Himalayas shows several islands of Area C, i.e., the least susceptible area, it is recommended that zone IV needs to be subdivided into smaller areas. This has tremendous implications in assigning design earthquake parameters and in planning and disaster mitigation at the micro level.

Area A overlaps the Kangra seismogenic source zone (SSZ) and Uttarakhand SSZ obtained in an earlier study (Mridula *et al.* 2014), shown in figure 11. Kangra SSZ was identified as the most vulnerable zone in terms of magnitude, return

periods, PGA and occurrence of earthquakes. When susceptible areas identified in this study were superimposed on SSZs, it was observed that most recent earthquakes were at the intersection of the Kangra SSZ and the most susceptible Area A.

Figure 12. Susceptible area map superimposed on the map of potential knots of $M_w \geq 6.5$ (after Bhatia *et al.* 1992).

Figure 13. Susceptible area map superimposed on the map of potential knots of $M_w \geq 7.0$ (after Bhatia *et al.* 1992).

These are located in Doda district of J&K and Chamba district of HP.

Bhatia *et al.*'s (1992) qualitative study for the occurrence of strong earthquakes identifies potential areas as dangerous knots in the Himalayan arc. The quantitative approach of the present study has identified three kinds of seismically susceptible areas in western Himalayas where Area A is the most susceptible. The susceptible area map was superimposed on the map of potential knots for earthquakes $M_w \geq 6.5$, identified by Bhatia *et al.* (1992), and compared as shown in figure 12. Several dangerous knots #10, 18, 19, 23, 24, 25, and 34 are located totally within Area A, and major parts of several other knots partially overlap Area A. Similarly, for map of potential knots for earthquakes $M_w \geq 7.0$, two knots #16 and #19, between 76°E and 78°E , and knot #23 between 78° and 79°E are completely within Area A, as shown in figure 13. Other knots also partially overlap some parts of Area A. Dangerous knots of Bhatia *et al.* (1992) are either within area A or are at boundary of A, which implies a similarity in identification of almost same areas by different methods. In addition, Area A is much larger than the area encompassed by knots. Gorshkov *et al.* (2012) validated the maps obtained by Bhatia *et al.* (1992) and found that within the past 20 years the Uttarkashi earthquake of 1991 occurred within knot #24, Chamoli earthquake of 1999 occurred within knot #35 and the Kashmir earthquake of 2005 occurred within knot #5. These earthquakes were part of the learning stage in the present study, and are incorporated in the MHD catalogue. These three nodes are within Area A of the present study. Two knots, i.e., #24 and 35 are within the largest Area A and #5 is in a different island of Area A.

4. Conclusions

This study is an attempt to identify seismically susceptible areas in the western Himalayas, using pattern recognition technique. Pattern recognition started with identification, selection and extraction of features from seismotectonic data. These features were then subjected to discriminant analysis and the study area was classified into three categories. The study was validated by testing earthquake data which occurred after completion of the MHD catalogue and was not part of the training exercise, i.e., for the years 2013 to 2015. It was found that four out of 70 earthquakes were of magnitude $M_w \geq 5.5$, and all four occurred in Area A, which has been assigned as the most susceptible area. Similar analysis was carried out for assigning moderately and least susceptible areas. Therefore, the study area was divided into three

susceptible areas, viz., Area A: most susceptible area, Area B: moderately susceptible area and Area C: least susceptible area. Results show that almost the entire states of Himachal Pradesh and Uttarakhand and a portion of Jammu & Kashmir are classified as Area A, while most of Jammu & Kashmir is classified as Area B and the Indo-Gangetic plains are classified as Area C.

By superimposing susceptible area map on seismic zoning map of India (IS: 1893 (Part 1)-2002), it was observed that most of the 70 earthquakes lie in the space between the three islands of zone V, i.e., in zone IV. It implies that zone V in BIS map needs modification in the western Himalayas and it should be considered as a continuous and larger area between and along the MBT and MCT, extending from J&K, HP and Uttarakhand. On similar lines as western Himalayas shows several islands of Area C, i.e., the least susceptible area, it is recommended that zone IV needs to be subdivided into smaller areas. This has tremendous implications in assigning design earthquake parameters and in planning and disaster mitigation at the micro level. The susceptible areas were also superimposed on the seismogenic source zones (SSZ) demarcated in a probabilistic seismic hazard assessment study (Mridula *et al.* 2014) and it was observed that of the 70 recent earthquakes, most were at intersection of Kangra SSZ, which was identified as most vulnerable SSZ and susceptible Area A. The susceptible area map was superimposed on the maps of potential knots for earthquakes $M_w \geq 6.5$ and $M_w \geq 7.0$, identified by Bhatia *et al.* (1992). The potentially dangerous knots identified by Bhatia *et al.* (1992) were part of the identified most susceptible Area A.

This study allowed us to identify and narrow down areas where most recent earthquakes are originating, and also where strong earthquakes may recur. These results can contribute to improve the characterization of seismic sources for seismic hazard assessment. Since the applied method is not intended to provide PGA, like by the DSHA and PSHA methods of hazard analysis; PGA estimates can be assigned to the identified susceptible areas by superimposing the PGA contours obtained by the probabilistic method (Mridula *et al.* 2014). Hypothetical earthquake scenarios can also be developed for the identified earthquake prone areas, as shown in Peresan *et al.* (2011).

These areas are under seismic threat and special mitigation measures must be taken to prevent any adverse consequences which may take place during any natural disaster specially earthquakes in these zones. These results emphasize that identified seismically susceptible areas that coincide with identified seismic source zones can be perceived as under

an enhanced threat perception. With concentration of dense population these are the areas in need of intervention and specific action for reducing the impact of an impending earthquake disaster. This may help in reducing losses in terms of life and infrastructure.

Acknowledgements

The authors would like to thank the Department of Earthquake Engineering, IIT Roorkee for providing academic support to carry out this research work. The authors thank the India Meteorological Department, New Delhi, for the epicentral data provided. Author (Mridula) is grateful to the Ministry of Human Resource Development (MHRD) for the financial support provided for the study.

References

- Bhatia S C, Chetty T R K, Filimonov M B, Gorshkov A I, Ya Rantsman E and Rao M N 1992 Identification of potential areas for the occurrence of strong earthquakes in Himalayan arc region; *Proc. Indian Acad. Sci.* **101** 369–385.
- Caputo M, Keilis-Borok V I, Oficerova E, Ranzman E, Rotwain I and Solovjeff A 1980 Pattern recognition of earthquake prone areas in Italy; *Phys. Earth Planet Interiors* **21** 305–320.
- Cartosat-1 data user's handbook, http://bhuvan.nrsc.gov.in/bhuvan_links.php.
- Dasgupta S, Mukhopadhyay B, Mukhopadhyay M and Nandy D R 2013 Role of transverse tectonics in the Himalayan collision: Further evidences from two contemporary earthquakes; *J. Geol. Soc. India* **81** 241–247.
- Davis J C 2002 *Statistics and Data Analysis in Geology*; 3rd edn, John Wiley and Sons, 638p.
- Duda R O, Hart P E and Stork D G 2001 *Pattern classification*; Wiley, New York.
- Gardner J K and Knopoff L 1974 Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?; *Bull. Seismol. Soc. Am.* **64** 1363–1367.
- Gelfand I M, Guberman Sh A, Izvekova M L, Keilis-Borok V I and Ia Ranzman E 1972 Criterion of high seismicity determined by pattern recognition; The upper mantle; *Tectonophysics* **13** 415–422.
- Gelfand I M, Guberman Sh A, Kaletskaia M S, Keilis-Borok V I, Ranzman E, Rotwain I M and Zhidkov M P 1974 Recognition of places where strong earthquakes may occur. II: Four regions of Asia Minor and S-E Europe (in Russian).
- Gelfand I M, Guberman Sh A, Keilis-Borok V, Knopoff L, Press F, Rantsman E, Rotwain I and Sadovsky A 1976 Pattern recognition applied to earthquake epicenters in California; *Phys. Earth Planet. Interiors* **11** 227–283.
- Gorshkov A I, Mokhtari M and Piotrovskaya E P 2009 The Alborz region: Identification of seismogenic nodes with morphostructural zoning and pattern recognition; *J. Seismol. Earthq. Eng.* **11** 1–16.
- Gorshkov A I, Soloviev A A, Jimenez M J, Fernandez G and Panza G F 2010 Recognition of earthquake prone areas ($M \geq 5.0$) in the Iberian peninsula; *Rend. Fis. Acc. Lincei* **21** 131–162.
- Gorshkov A, Parvez I A and Novikova O 2012 Recognition of earthquake-prone areas in the Himalaya: Validity of the results; *Int. J. Geophys.*, Article ID 419143, 5p, doi: 10.1155/2012/419143.
- Gupta I and Sinvhal A 2010 Assessment of seismic risk in a micro zone; *Proc. 14th Symposium on Earthquake Engineering*, pp. 224–232.
- Gupta I, Sinvhal A and Shankar R 2006 Himalayan population at risk: Strategies for preparedness; *Dis. Prev. Manag.* **5** 608–619.
- IS: 1893 (Part 1)-2002 Earthquake hazard zoning map of India; www.bis.org.in.
- ISC Catalogue: <http://www.isc.ac.uk/iscbulletin/search/bulletin>.
- King G 1986 Speculations on the geometry of the initiation and termination processes of earthquake rupture and its relation to morphology and geological structure; *Pure Appl. Geophys.* **124** 567–583.
- Mridula, Sinvhal A and Wason H R 2013 A review on pattern recognition techniques for seismic hazard analysis; *Proc. Int. Conf. Emerging Trends in Eng. and Tech (ACEEE)*, pp. 854–858.
- Mridula, Sinvhal A and Wason H R 2014 Probabilistic seismic hazard assessment in the vicinity of MBT and MCT in western Himalaya; *Research Inventory: Int. J. Eng. Sci.* **4(11)** 21–34.
- Mridula, Sinvhal A, Wason H R and Rajput S S 2016 Segmentation of Main Boundary Thrust and Main Central Thrust in western Himalaya for assessment of seismic hazard, communicated.
- Mukhopadhyay B 2011 Clusters of moderate size earthquakes along Main Central Thrust (MCT) in Himalaya; *Int. J. Geosci.* **2** 318–325, doi: 10.4236/ijg.2011.23034.
- Mukhopadhyay B, Acharyya A and Dasgupta S 2011 Potential source zones for Himalayan earthquakes: Constraints from spatial-temporal clusters; *Nat. Hazards* **57** 369–383, doi: 10.1007/s11069-010-9618-2.
- Narula P L, Acharyya S K and Banerjee J 2000 *Seismotectonic Atlas of India and its Environs*; Geological Survey of India, New Delhi.
- Peresan A, Zuccolo E, Vaccari F, Gorshkov A and Panza G F 2011 Neo-deterministic seismic hazard and pattern recognition techniques: Time-dependent scenarios for north-eastern Italy; *Pure Appl. Geophys.* **168** 583–607.
- Peresan A, Gorshkov A, Soloviev A and Panza G F 2015 The contribution of pattern recognition of seismic and morphostructural data to seismic hazard assessment; *Bull. Geof. Theor. Appl.* **56** 295–328.
- Ristau J, Rogers G C and Cassidy G F 2005 Moment magnitude–local magnitude calibration for earthquakes in western Canada; *Bull. Seismol. Soc. Am.* **95** 1994–2000.
- Sinvhal A 2010 Understanding earthquake disasters; Tata McGraw Hill Education Pvt. Ltd., p. 285.
- Sinvhal A, Joshi G, Sinvhal H and Singh V N 1990 A pattern recognition technique for microzonation; *Proc. 9th Symp. Earthq. Eng.* **1** 24–30.
- Sinvhal A, Sinvhal H and Joshi G 1991 A valid pattern of microzonation; *Proc. 4th International Conference on Seismic Zonation III* 641–648.
- Stepp J C 1972 Analysis of completeness in the earthquake sample in the Puget Sound Area and its effect on statistical estimates of earthquake hazard; *Proc. Int. Conf. Microzonation*, Seattle, Washington.

- Talwani P 1988 The intersection model for intraplate earthquakes; *Seismol. Res. Lett.* **59** 305–310.
- Talwani P 1999 Fault geometry and earthquakes in continental interiors; *Tectonophys.* **305** 371–379.
- USGS Catalogue: <http://earthquake.usgs.gov/earthquakes/eqarchives>.
- Varunoday, Gaur V K and Wason H R 1979 Spatial prediction of earthquakes in the Kumaon Himalaya by pattern recognition; *Mausam* **30** 253–264.
- Walt C M and Barnard E 2007 Data characteristics that determine classifier performance; *SAIEE Africa Res. J.* **98(3)** 87–93.
- Wason H R, Das R and Sharma M L 2012 Magnitude conversion problem using general orthogonal regression; *Geophys. J. Int.* **190** 1091–1096.
- Wiemer S and Wyss M 2000 Minimum magnitude of complete reporting in earthquake catalogs: Examples from Alaska, the western United States and Japan; *Bull. Seismol. Soc. Am.* **90** 859–869.
- Wyss M, Hasegawa A, Wiemer S and Umino N 1999 Quantitative mapping of precursory seismic quiescence before the 1989, m7.1 off-Sanriku earthquake, Japan; *Ann. Geophys.* **42(5)** 851–869.

MS received 9 June 2015; revised 22 January 2016; accepted 1 February 2016

Corresponding editor: PAWAN DEWANGAN