On intermediate-term prediction of strong earthquakes in the Himalayan arc region using pattern recognition algorithm M8

S C BHATIA, S V CHALAM, V K GAUR, V I KEILIS-BOROK* and V G KOSOBOKOV*

National Geophysical Research Institute, Hyderabad 500 007, India
*Institute of Physics of the Earth, Moscow, USSR

Abstract. Seismicity of the Himalayan arc lying within the limits shown in figure 1 and covering the period 1964 to 1987 was scanned using M8 algorithm with a view to identifying the times of increased probabilities (TIPs) of the occurrence of earthquakes of magnitude greater than or equal to 7·0, during the period 1970 to 1987. In this period, TIPs occupy 18% of the space time considered. One of these precedes the only earthquake in this magnitude range which occurred during the period. Two numerical parameters used in the algorithm, namely the magnitude thresholds, had to be altered for the present study owing to incomplete data. Further monitoring of TIPs is however warranted, both for testing the predictive capability of this algorithm in the Himalayan region and for creating a base for the search of short-term precursors.

Keywords. Himalaya; intermediate term prediction; M8 algorithm; pattern recognition algorithm; times of increased probabilities; seismicity.

1. Introduction

In any scheme of pattern recognition, distinctive patterns of attributes describing the traits of events (objects) of a particular class are first established in the course of a learning process. The ensemble of objects which are thus unequivocally grouped into two or more classes and the attributes describing them, constitute the learning material from which recognition rules are abstracted. These rules are then applied to recognition of new objects, that is assigning them to the appropriate class, which is the ultimate objective. For predicting earthquakes, the objective is to recognize instants of times which presage the approach of a strong earthquake. This can be attempted by establishing premonitory patterns based on the analysis of past data and designing an appropriate pattern recognition algorithm.

Earthquake sequences constitute so far the longest and the least incomplete data set relevant to earthquake prediction in many parts of the world. This is also largely true for the Himalayan region. Here we analyse the earthquake sequences in the Himalayan region using the M8 algorithm designed by Keilis-Borok and Kosobokov (1984) to test the possibility of diagnosing the approach of a strong earthquake. The premonitory patterns expressed by this algorithm appear to be rather similar in widely different regions of the world (Gabrielov et al 1986).

It is of course recognized, at the very outset, that the symptoms of an approaching earthquake can be quite different from one case to another. This may happen because
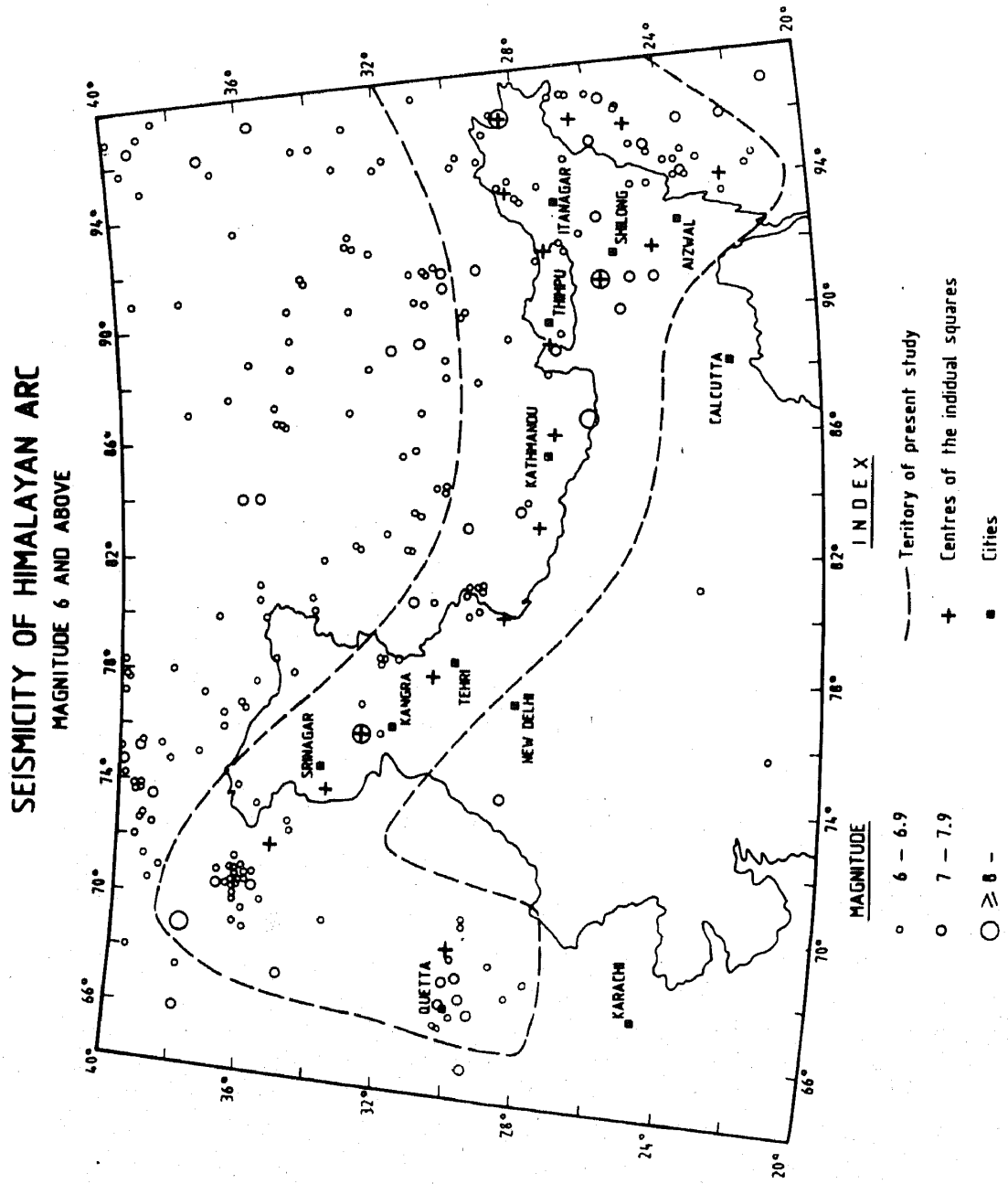
Figure 1.   Seismicity map of Himalayas and adjacent regions showing earthquakes of magnitude 6 and above and the territory scanned in present study.

of the distinctive seismotectonic environment of different seismic regions or due to inhomogeneous distribution of rock materials and stresses within the same region. Even earthquakes occurring in the same area may show different premonitory associations if the stress regime and rock properties preparatory to fracture have been greatly altered by earlier or neighbouring earthquakes. The nature of earthquake sequences is therefore so diverse (Gabrielov *et al* 1987), as to defy classification and frustrate attempts to abstract from them a consistent set of premonitory signals. In order to circumvent this problem, Allen *et al* (1984) and Gabrielov *et al* (1986) devised a scheme of integral representation of earthquake sequences which smoothens the diversity of seismicity patterns without, however, losing any premonitory signals that may be contained in the data. This is realized in the following way.

First, the earthquake flow within a certain large territory is carefully scanned. Several integral functions (attributes) describing the nature (traits) of the earthquake sequence in the region are then evaluated within a sliding time window, each assigned to the instant of time that terminates the window. In all, 14 functions describing the following traits can be formed: level of seismicity, seismic quiescence, variation of seismicity, spatial concentration, clustering in time and space, contrast of seismicity in adjacent areas and long term interaction.

Figure 2 shows a schematic representation of traits and the manner of their evaluation along the time axis. Appropriate pattern recognition algorithms can be developed using some or all of these traits. Diagnosis of the time of increased probability of occurrence of a strong earthquake (TIP) is made by applying the appropriate recognition rules. After a TIP has been diagnosed, a strong earthquake of magnitude greater than or equal to a threshold magnitude M_0 is expected to occur within a few hundred kilometers over the next few years. This is called an intermediate-term prediction. A magnitude threshold M_0 is determined for each region on the basis of its past history of strong earthquakes.

To ensure uniform diagnostics of TIPs for different regions, the functions describing various traits are normalized, as described below. The sequence of main shocks is first isolated, the number of aftershocks being considered as one of the characteristic functions of the corresponding main shock. In order to smoothen the earthquake sequence, their traits are first defined within a wide time window and epicentral area. Next, groups are formed corresponding to each of the functions, which comprise a set of these functions evaluated for different arguments. For example, the group of the function describing the intensity of earthquake flow consists of the values of the function evaluated for different magnitude ranges. Each function is then quantized in two or three discrete intervals, and various combinations of these are used to diagnose a TIP.

Two related algorithms for such diagnosis of TIPs are available; one called CN, designed by Allen *et al* (1984) and the other, M8 designed by Keilis-Borok and Kosobokov (1984). Their application to different regions of the world is described by Gabrielov *et al* (1986). Algorithm M8, used in the present study, is described below.

2. Pattern recognition algorithm M8

The algorithm M8, uses the following traits described by suitable numerical functions: the number of main shocks, its deviation from long term trend, spatial concentration
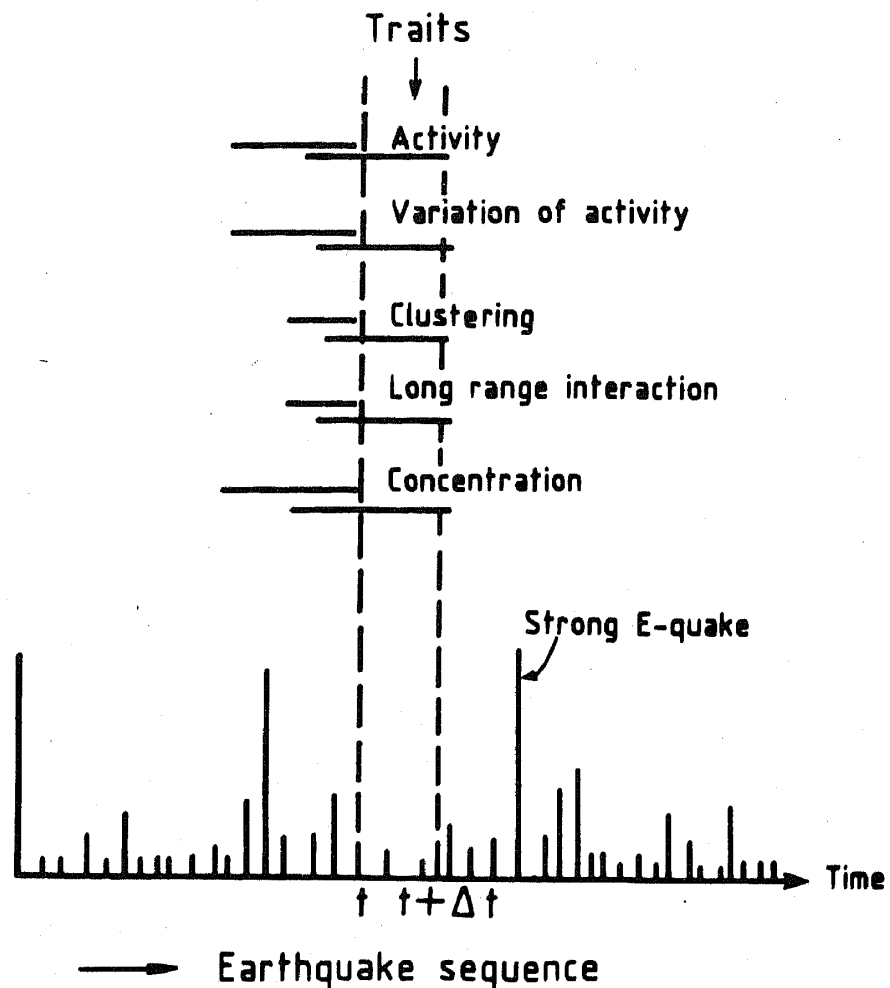
Figure 2. Schematic representation of the sequence of earthquake occurrence and evaluation of intrinsic traits for PR methods. Vertical lines on the time axis show the occurrence of earthquakes (the length being proportional to the magnitude). Various traits are evaluated along sliding time window at Δt steps. Broken vertical lines show this at t and $t + \Delta t$ time instants.

of sources and clustering of sources, measured by the maximal number of aftershocks of the main shocks. The territory under study is scanned by a network of overlapping squares. These squares are centred at appropriate points which show comparatively greater seismic activity with respect to their surroundings. The side length (L) of each square in degrees of the great circle is determined by the relation

$$L(M_0) = \exp(M_0 - 5 \cdot 6) + 1.$$

The above functions are evaluated from the data of earthquake flow in these squares. Each of the first three functions is determined twice, for two different magnitude thresholds M_1 and M_2. The set of values of a function evaluated for different arguments together constitute a 'group'. The values M_1 and M_2 are determined by the criterion that the average annual number of such main shocks is greater than or equal to the predetermined numbers n_1 and n_2 respectively. These values of n_1 and n_2 are consistently adopted for all squares within a region and for all regions, in order to normalize the seismic activity. In defining the spatial concentration, the main shocks

to be considered, lie within the magnitude range M_1 to $(M_0 - 0.2)$ while for clustering they lie in the range $(M_0 - 2.0)$ to $(M_0 - 0.1)$. The flow of main shocks is thus normalized with respect to M_0. The time window for which the functions are evaluated is 6 years for the first three functions and one year for clustering. The sliding step is chosen appropriately, say 6 months.

A function is identified as being abnormally large when its value exceeds the prescribed value, say the 10% percentile of the range of values of the function. Pattern M8, is said to have occurred if at two successive instants of time $K - 1$ and K, at least a prescribed number H of individual functions attain abnormally large values, and there are at least a prescribed number G of groups which have at least one member that attains an abnormally large value. A TIP is then declared at the time instant K. The duration of TIP is a suitable multiple (say 10) of the sliding time step. The free parameters of algorithm M8 were determined by Keilis-Borok and Kosobokov (1984) by establishing TIP diagnosis in respect of the strongest earthquakes of magnitude 8 or greater that occurred between 1970 and 1983 in different parts of the earth. The algorithm was later successfully applied to other regions as well, with minimal readaptation of parameters; only the magnitude threshold M_0 had to be adjusted for every region (Gabrielov *et al* 1986). For example, M_0 was 8 for Mexico and 6 for Western Canada. The success-to-failure score was no worse than in the original data-fitting (Keilis-Borok and Kosobokov 1984). Similar results were obtained in tests for algorithm CN (Gabrielov *et al* 1986) with a very low value of M_0 equal to 4.5.

However, as this test is retrospective, it cannot be considered to be conclusive. But it does support the concept of global self-similarity of earthquake flow (Sadovsky *et al* 1982; Gabrielov *et al* 1986) and warrants similar tests to be made for other regions.

3. The region

The region considered for this study is the Himalayan sector (figure 1) of the Alpide-Himalayan fold belt, including the Indus-Ganga-Brahmaputra foredeep to the south of the ranges and part of the Burmese arc. The evolution of the Himalaya is traced to the collision of the Indian plate with the Eurasian plate. Its present-day seismicity is attributed to continued convergence of the Indian plate towards the north by underthrusting beneath the Tibetan plateau. Owing to the gentle dip of this underthrusting plane which leads to a wide zone of frictional contact, the seismic activity along the belt is quite diffuse and mostly confined within a depth of about 20 km. At its eastern extremity however, along the Arakan-Yoma fold belt, hypocentral depths range from shallow to intermediate (about 200 km). Although major earthquakes periodically occur all along the Himalayan collision boundary to relieve the accumulating strains, some parts of the belt show more prolific seismicity, notable among which is the eastern segment.

Four devastating earthquakes of magnitude greater than 8.4 have ruptured the various segments of the 2400 km long Himalayan arc during the past 100 years, leaving seismic gaps of different lengths in between. Unfortunately, our knowledge about the seismicity of the Himalaya in preceding centuries is very sketchy although no earthquakes of comparable magnitude are known to have occurred since the

sixteenth century. It is therefore difficult to estimate the recurrence period of great earthquakes in the region, although earthquakes of magnitudes less than 5 are quite frequent, interspersed occasionally by earthquakes of magnitude 5 to 6. Since 1970 however, the entire region seems to have entered a period of quiescence as far as earthquakes of magnitude 7 or greater are concerned.

4. The data

The earthquake catalogue selected for the present study was taken from the NOAA World Hypocenter Data File issue of 1987. More detailed catalogues based on local networks are still restricted to small areas and limited time-intervals. A more complete global catalogue is of course produced by the International Seismological Centre at Newbury, England, but the long delay in dissemination of this catalogue makes it unsuitable for this study.

The main shocks were isolated from the aftershocks by the algorithm described by Keilis-Borok *et al* (1980) using the space and time windows given in table 1.

Epicentral locations of the main shocks are given in tables 2 and 3. The territory lying between latitudes 21·0°N and 36·5°N and longitudes 65·6°E–99·0°E was carefully scanned (figure 1). It is framed by the Hindu Kush mountains in the northwest famous for their strong seismicity at intermediate depths, and by the parameridional Sichwan seismic belt lying between 102°E and 105°E.

The temporal chart of the main shocks given in table 4 depicts a magnitude-wise distribution of earthquakes in the various rectangular areas bounded by latitudes 17°N–45°N and longitudes 60°E–105°E. It may be recalled that algorithm M8 can only be applied to diagnose TIPs for the period which starts 7 years after the first availability of a sufficiently complete catalogue, as some of the functions are evaluated from data within a time window of six years prior to the instant of their evaluation. Also, the data pertaining to the first 6 years are necessary to establish as to which values of the functions are abnormally large. Table 4 shows that the catalogue is relatively complete only for magnitudes greater than 4·5–5·0, and that too after 1964.

Table 1. Windows for the identification of after-shocks (reproduced from Keilis-Borok *et al* 1980).

M	R (km)	T (days)
2·5	19·5	6·0
3·0	22·5	11·5
3·5	26·0	22·0
4·0	30·0	42·0
4·5	35·0	83·0
5·0	40·0	155·0
5·5	47·0	290·0
6·0	54·0	510·0
6·5	61·0	790·0
7·0	70·0	915·0
7·5	81·0	960·0
8·0	94·0	985·0

Table 2. Magnitude of the largest earthquake in each of the zones defined by the latitude and longitude.

Latitude (N)	Longitude (E) 60·0	62·5	65·0	67·5	70·0	72·5	75·0	77·5	80·0	82·5	85·0	87·5	90·0	92·5	95·0	97·5	100·0	102·5	105·0
45·0	—	—	0·0	—	—	0·0	4·8	8·7	6·5	7·2	8·3	6·6	7·3	7·5	4·6	4·8	6·5	5·6	—
43·0	0·0	4·3	6·5	5·5	5·5	7·6	6·8	7·1	7·3	7·2	6·8	5·6	4·7	5·3	4·8	4·7	—	7·0	—
41·0	4·9	7·0	4·7	7·1	6·1	7·8	8·6	6·8	5·1	7·3	4·6	4·9	6·5	—	7·6	4·0	—	5·0	—
39·0	4·9	5·1	7·2	8·1	7·6	7·0	6·4	6·0	5·0	5·8	6·8	5·5	6·3	4·6	7·2	6·5	7·2	7·0	—
37·0	4·5	3·9	5·6	6·6	8·1	6·8	6·8	6·3	6·3	7·2	6·9	6·0	5·0	6·2	5·5	7·6	8·3	6·8	—
35·0	5·3	—	5·8	7·6	5·8	6·0	5·6	6·5	6·5	5·2	5·5	6·0	6·5	6·1	6·5	4·9	6·0	5·5	—
33·0	6·3	0·0	5·6	5·3	4·9	5·2	8·6	7·1	6·5	6·8	6·3	7·1	8·0	6·0	5·8	7·7	7·4	7·4	—
31·0	5·4	4·8	7·5	7·0	6·1	4·6	4·2	5·5	7·7	7·3	5·8	6·2	7·9	5·0	6·0	6·5	7·4	6·0	—
29·0	6·7	7·0	5·4	6·8	0·0	7·1	4·7	5·6	4·7	7·0	5·7	6·3	6·8	7·9	8·7	6·3	6·8	7·5	—
27·0	6·5	5·2	5·9	5·6	5·0	—	0·0	—	0·0	—	8·4	5·5	7·1	7·2	7·6	6·5	7·3	6·0	—
25·0	4·7	8·3	5·6	5·2	5·6	—	—	0·0	6·5	—	4·8	6·3	7·6	7·2	7·8	7·0	7·2	7·5	—
23·0	5·5	0·0	0·0	0·0	—	5·4	6·3	4·9	5·5	0·0	0·0	5·5	5·2	6·5	7·0	7·3	7·0	6·6	—
21·0	4·8	0·0	—	5·2	—	4·5	—	—	—	—	4·4	—	5·2	5·5	7·9	5·5	6·8	6·8	—
19·0	5·6	—	—	—	0·0	6·0	—	0·0	5·7	—	5·0	—	—	5·6	5·6	5·9	—	4·7	—
17·0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	7·3	—	—	—	—

Table 3. Number of earthquake in each of the zones defined by the latitude and longitude.

Latitude (N)	Longitude (E) 60.0	62.5	65.0	67.5	70.0	72.5	75.0	77.5	80.0	82.5	85.0	87.5	90.0	92.5	95.0	97.5	100.0	102.5
45.0	—	1	—	—	—	2	5	20	19	30	29	14	5	6	2	6	14	9
	1	7	2	15	31	29	42	40	29	32	21	14	4	5	1	3	—	2
41.0	8	6	33	34	144	146	91	97	5	3	1	6	6	—	9	4	—	5
	3	16	2	125	364	197	54	13	6	1	8	5	12	3	14	13	17	8
37.0	1	13	4	180	500	50	33	40	28	8	12	7	4	6	13	4	6	6
	9	2	—	34	38	50	36	15	34	10	7	19	15	11	13	8	11	12
33.0	8	10	2	26	22	18	42	30	10	19	22	26	17	26	21	7	12	28
	5	29	3	85	36	3	1	33	67	20	18	17	21	21	39	22	23	23
29.0	16	36	6	23	2	1	3	2	10	23	35	21	30	31	50	16	29	28
	26	28	12	12	3	—	1	—	1	—	3	12	60	37	128	46	26	18
25.0	3	11	21	12	9	1	—	3	2	—	1	1	23	142	61	16	17	16
	5	1	1	1	—	3	1	2	1	1	2	2	5	95	20	16	29	5
21.0	5	1	1	1	—	8	—	—	—	—	1	—	1	15	14	13	15	3
	4	—	—	—	1	8	—	1	4	—	2	—	—	14	17	3	—	1
17.0	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—

Table 4. Number of earthquakes in the magnitude ranges and time windows in the region defined by latitude 17·0°N to 45·0°N and longitude 60·0°E to 105·0°E. The numbers in the first column correspond to earthquakes with unknown magnitudes.

Year	0·0	1·0	2·0	3·0		4·0		5·0		6·0		7·0		8·0		9·0
1900	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—	1
	3	—	—	—	—	—	—	—	—	—	1	6	2	1	—	1
1910	7	—	—	—	—	—	—	—	—	2	1	3	3	1	—	1
	17	—	—	—	—	—	—	—	—	2	3	2	—	—	—	—
1920	56	—	—	—	—	—	—	—	1	2	5	2	2	1	—	1
	159	—	—	—	—	—	—	—	7	7	8	2	2	—	—	—
1930	128	—	—	—	—	—	—	—	30	10	5	4	—	—	—	—
	146	—	—	—	—	—	—	—	8	14	2	3	5	—	—	—
1940	60	—	—	—	—	—	—	—	8	6	4	6	—	—	—	—
	12	—	—	—	—	—	—	—	1	2	2	—	—	—	—	—
1950	336	—	—	—	—	—	—	—	1	10	4	3	—	—	1	1
	184	—	—	—	—	—	—	6	4	13	6	3	—	—	—	—
1960	285	—	—	2	2	9	37	32	15	13	8	2	1	—	—	1
	238	—	—	6	1	14	130	130	33	11	3	—	—	—	—	—
1970	226	—	—	—	7	26	134	137	30	10	3	—	—	—	—	—
	391	—	—	—	24	101	246	117	12	9	4	4	1	—	—	1
1980	384	—	—	1	14	101	347	91	15	7	6	1	—	—	—	—
	279	—	—	—	9	99	185	36	7	3	2	—	—	—	—	—
1990	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

The diagnosis of TIPs can therefore be attempted only for the period commencing from 1970.

Additionally, we have also considered an isolated cluster of epicenters in the Quetta region (around 30°N, 68°E). A temporal chart for this region is given in table 5.

5. Selection of M_0

Four earthquake with M greater than 8 and a larger number with magnitudes ranging between 7 and 8 have occurred during the past 100 years in the Himalayan region, which is the area chosen for the present study (table 4). Considering the 2400 km length of the Himalayan collision zone, an equal number of such earthquakes could possibly occur in the future in other segments of the Himalayan belt, particularly in the region flanked by the rupture zones of these 4 earthquakes. However, since about 1970, the region has been relatively quiet. During 1970–1987, only one earthquake of magnitude 7·0 or a little higher has taken place. One can, therefore, test algorithm M8 only for M_0 greater than or equal to 7. For larger M_0, only false alarms, but not the failure to predict, could be considered.

6. Diagnosis of TIPs

The area under study was spanned by 18 squares centred at points shown in figure 3. For $M_0 = 7$, the length (L) of the side of squares turns out to be 560 km, according to the formula stated earlier. The lower magnitude cut-off in the catalogue is, however, too high (5·0). Therefore, the average annual number of the main shocks is well below the required number 20. In some of the squares, this figure is as small as 3. Therefore, values of the threshold n had to be lowered. For the common values in all squares, we assumed $n_1 = 3$ and $n_2 = 1$, instead of 20 and 10, normally used in earlier studies. The other possible alternative of using $n_2 = 2$ was also tried.

The threshold values of G and H chosen for recognizing the instant of TIPs, were 4 and 6 respectively, the same as used in other parts of the world where M8 has been applied. Application of this algorithm to the Himalayan belt yeilded TIPs preceding strong earthquakes, in the square covering the India-Burma border region, as well as some false alarms in others. These results are shown in figure 3. They are reasonably acceptable and practically the same for both values of n_2. The TIPs occupy 18% of the total time-space, and precede the only strong earthquake ($M_b = 7·1$) that occurred on 29 May 1976, in the India-Burma border region (24·53°N 98·71°E). A TIP is also recognized in the Quetta region preceeding a strong earthquake of magnitude 7·0 which occurred there (33·025°N, 66·315°E) on 3 October 1975. Some agencies reported this earthquake as having a magnitude of 6·7.

7. Discussion of results

In applying M8 to the present study, the criterion used for recognition of TIPs was the same as for other regions of the world. However, the parameters n_1 and n_2 were adjusted in the manner explained above for the purpose of defining magnitude
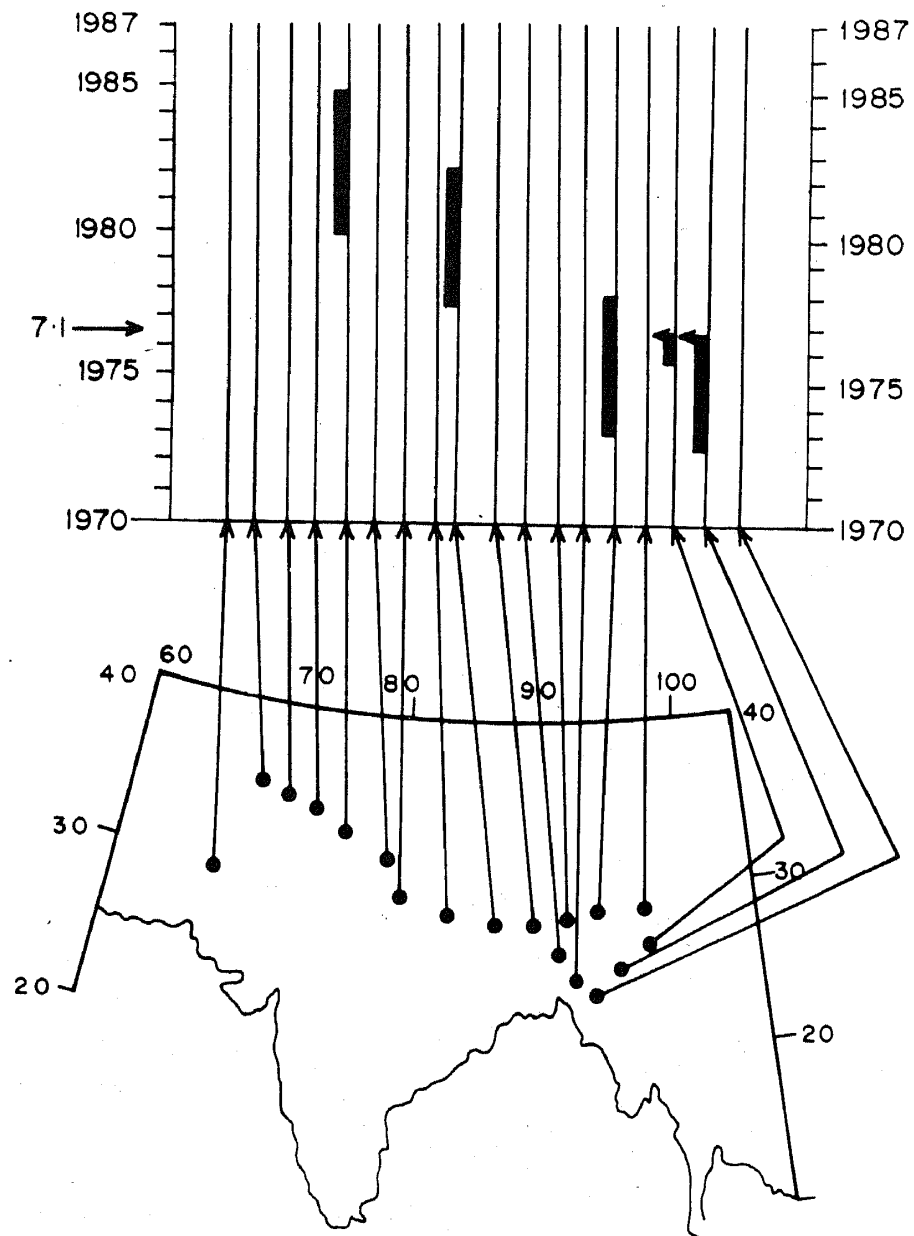
Table 5. Number of earthquakes in the magnitude ranges and time windows in the Quetta region defined by latitude 29·0°N to 31·0°N and longitude 61·0°E to 63·0°E. The numbers in the first column correspond to earthquakes with unknown magnitudes.

Magnitude Year	0·0	1·0	2·0	3·0	3·5	4·0	4·5	5·0	5·5	6·0	7·0	8·0	9·0
1900	—												
1910	—										1		
1920	2									1			
1930	1									1	2		
1940	1												
1950	5												
1960	2						1	1	1	1			
1970	3				5	2	4	3	1	1			
1980						3	2	1	1				
1990						8	5	5					
						2	10	1					
							4	3					

Figure 3. Big dots on the map show the centres of the overlapping squares studied using M8 algorithm. Darkened segments along the time axis corresponding to each of the dots indicate TIPs in that square. TIPs terminated by an arrow indicate occurrence of the predicted event.

thresholds. This enforced change in the values of n_1 and n_2, makes it difficult to directly support our results by those of earlier studies in other regions. On the other hand, this change hardly made the diagnosis easier. When n is small, the random variations of all functions are expected to increase, thereby increasing the errors. However, our experience in carrying out this exercise on Himalayan data did not bring out such randomness. This, though contrary to expectation, is nevertheless encouraging. For, this may be the result of a higher cut-off for lower magnitude thresholds of detectability and a near completeness of the catalogue. The analysis was repeated by adopting 1967 as the cut-off data as the catalogue of lower magnitude

earthquake that occurred after that date was expected to be more complete. This exercise however, yielded only marginally better results.

These results indicate that future monitoring of TIPs in the Himalayan region, using this algorithm may prove to be quite illuminating. It may however be necessary to develop region-specific parameter thresholds for various segments of the Himalayan belt. To do this, the available seismic data recorded by the national network of stations over the past half century need to be rigorously studied to produce a more representative and complete catalogue by including even such events which are not so well located.

Acknowledgements

References

Allen C, Hutton K, Keilis-Borok V I, Kuznetsov I V and Rotvain I M 1984 Dolgosrochni prognoz Zemletryaseni i avtomodelnost seismologicheskih predvestinikov. Kalifornia, (Long-term earthquakes' prediction and self-similarity of premonitory seismicity patterns. California, M>/6·4, M>/7, in Russian). In *Dostizheniya i problemi sovermennoi geofisici* (Computational Seismology issue 21) (Nauka: Moskva) pp. 152–165

Gabrielov A M, Dimitrieava O E, Keilis-Borok V I, Kosobokov V G, Kuznetsov I V, Levshina T A, Mirzoev K M, Molchan G M, Negmatullaev S Kh, Pisarenko V P, Prozeroff A G, Rinehart W, Rotvain I M, Shebalin P N, Shnirman M G and Schreider S Yu 1986 Algorithms of long term earthquakes' prediction; *Rev. Geofis.* (in press) Historia (Comision de Geofisica)

Gabrielov A M, Keilis-Borok V I, Levshina T A and Shaphoshinkov V A 1987 Block model of the dynamics of the lithosphere; in *Computational seismology* (New York: Allerton Press) issue 19 (translation from Russian)

Keilis-Borok V I and Kosobokov V G 1984 A set of premonitory seismicity patterns before strongest earthquakes of the world. In *Earthquakes and mitigation of natural hazards, XXVII International Geological Congress Colloquium 06.* (Nauka: Moskva) Vol. 61, 56–66

Keilis-Borok V I, Knopoff L, Rotvain I M and Sidorenko T M 1980 Bursts of seismicity as long-term precursors of strong earthquakes; *J. Geophys. Res.* **85** 803–811

Kosobokov V G 1982 Experience in transferring criteria of high seismicity (M>/8·2) from the Circuve—Pacific belt to the Alpine belt; *Computational seismology* (New York: Allerton Press) issue 13 (translation from Russian)

Sadovsky M A, Bolchovitinov L G and Pisarenko V F 1982 On discreteness in solid earth material; Izvestiya Akademii Nauk USSR, Fizika Zemli N 12 pp 3–18