

Application of pattern recognition algorithm in the seismic belts of Indian convergent plate margins – M8 algorithm

V K GAHALAUT^{1,2}, I V KUZNETSOV^{1,3}, V G KOSOBOKOV^{1,3},
A M GABRIELOV^{1,3} and V I KEILIS-BOROK^{1,3}

¹International Institute for Earth, Environmental and Marine Sciences (IIEEM), P.O. Box 586, 34100 Trieste, Italy

²Department of Earth Sciences, University of Roorkee, Roorkee 247 667, U.P., India

³International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Academy of Sciences of the USSR, Warshavskoye sh. 79, K.2, Moscow 113 556, U.S.S.R.

MS received 27 April 1992; revised 21 July 1992

Abstract. The seismicity associated with the convergence of the Indian and Eurasian plates, from 1964 to August 1990 was scanned using the M8 algorithm with a view to identify the times of increased probabilities (TIPs) of the earthquakes of magnitudes greater or equal to 6.4 that occurred during the period from 1970 to August 1990. 23 out of 28 earthquakes ($M \geq 6.4$) have been predicted. These were preceded by specific activation of the earthquake flow which was picked up by the M8 algorithm. The earthquake of August 1988 in the Himalaya could not be predicted, the other four unpredictable earthquakes occurred in the early dates of the catalogue (1970–1971) and hence their TIPs could not be diagnosed. Two current alarms are diagnosed, one in the Indo-Burmese arc and the other in the Hindukush-Pamir region.

The algorithm provides the correlation between the earthquakes and their area of activation (both in time as well as in space) which, when compared with the local geology, may help to comment on the present day status of the seismic features on the surface.

Keywords. Intermediate term prediction; M8 algorithm; pattern recognition; times of increased probability.

1. Introduction

The Himalayan mountain chain runs unbroken for a distance of about 2400 kms and joins the Indo-Burmese mountains in the east and Hindukush-Pamir in the northwest. This chain is inhabited by many earthquakes due to the process of underthrusting of the Indian plate beneath the Eurasian plate (figure 1). There are atleast four major earthquakes ($M \geq 8$) which occurred during the past 100 years. To mitigate the seismic risk due to future earthquakes of this seismic belt, there is a need to develop a prediction method from the available data.

After the successful application of algorithm M8 in different seismic regions of the world (Keilis-Borok and Kosobokov 1990 and Gabrielov *et al* 1986), an attempt has been made to apply the algorithm in the seismic belts of the Indian convergent plate margin with the hope of diagnosing precursors which may help to predict the next pending earthquake ($M \geq 6.4$, magnitude here referred is common maximum) in the region. For this purpose, the NEIC catalogue from 1964 to August 1990 is used to find out the TIPs before the earthquakes ($M \geq 6.4$) which occurred after 1970 (table 1). The aftershocks were removed using the time and space windows given by

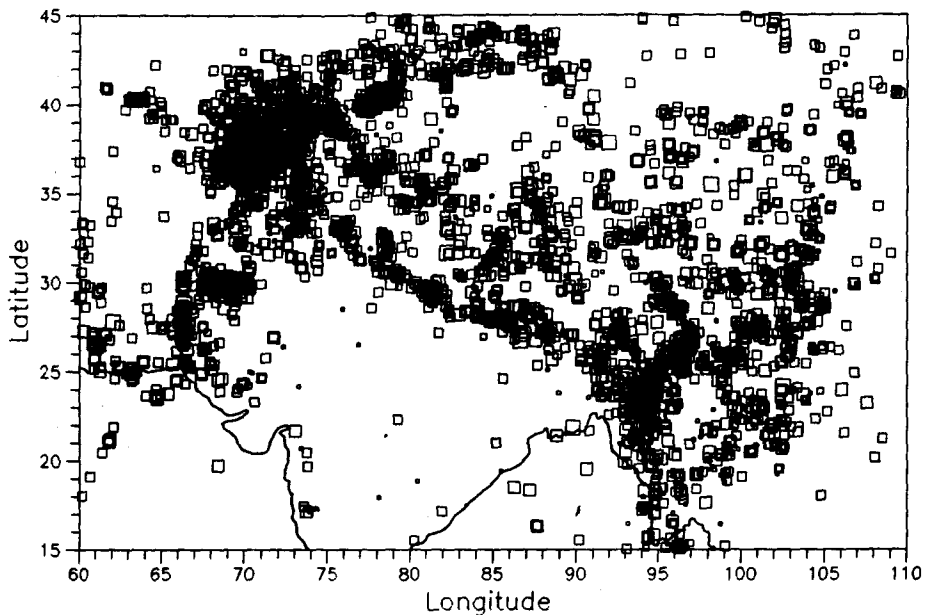


Figure 1. The main shocks from 1970 to August 1990 are plotted here. The minimum size of the square refers to the earthquakes of magnitude 0–3. The increase in the size of square corresponds to the increase in magnitude by one unit.

Keilis-Borok *et al* (1980). For the terminology of the faults and their geographic location, we referred Tapponier and Molnar (1977).

2. Algorithm M8

Before the occurrence of the event the distinctive patterns are established using the past data, which describes the traits of the event and are called the objects. The ensemble of objects which are grouped into two or more classes constitute the learning material from which recognition rules are abstracted. These rules are then applied to the recognition of new objects, which is the general scheme of pattern recognition algorithm M8.

The seismicity and the frequency of major earthquake occurrences is not the same in different regions. But it can be normalized choosing the annual rate of seismicity (n_1 and n_2) and the magnitude threshold (M_0) for a particular region. Thus the algorithm can be applied in any seismic region of the world without changing the other parameters. This is what has been achieved in the algorithm M8.

Here we define some functions which includes the intensity of an earthquake flow (N) or level of seismic activity, the deviation of seismic activity from a long linear trend (L), the concentration of mainshocks in space (Z) and the clustering of earthquakes by the maximal count of aftershocks (which is one of the parameters of a mainshock) in the last year (B).

As the intensity of earthquake flow in the regions is different, we normalize it by adjusting the magnitude threshold \underline{M} so that the average yearly rate of occurrence

Table 1. The time, location, depth and the four magnitudes of the strong mainshocks (from the NEIC catalogue) are given in the table, occurred between 15 and 45° latitude and 60 and 110° longitude. The events marked with * lie outside the regions where M8 is not applied.

| Year | Month | Day | Hour | Minute | Lat | Long | Depth | m_b | m_s | m_l | m_p |
|-------|-------|-----|------|--------|-------|--------|-------|-------|-------|-------|-------|
| 1970 | 1 | 4 | 17 | 0 | 24-14 | 102-50 | 31 | 5-90 | 7-50 | 0-00 | 7-50 |
| 1970 | 6 | 5 | 4 | 53 | 42-48 | 78-76 | 20 | 6-00 | 6-60 | 0-00 | 6-50 |
| 1970 | 7 | 29 | 10 | 16 | 26-02 | 95-40 | 59 | 6-50 | 0-00 | 0-00 | 6-40 |
| 1971 | 3 | 23 | 9 | 52 | 41-49 | 79-26 | 0 | 5-70 | 0-00 | 0-00 | 6-80 |
| 1971 | 3 | 23 | 20 | 47 | 41-45 | 79-26 | 0 | 6-00 | 5-80 | 0-00 | 6-90 |
| *1971 | 3 | 24 | 13 | 54 | 35-46 | 98-17 | 13 | 5-80 | 6-40 | 0-00 | 0-00 |
| 1973 | 2 | 6 | 10 | 37 | 31-40 | 100-58 | 0 | 6-10 | 7-40 | 0-00 | 7-70 |
| *1973 | 7 | 14 | 4 | 51 | 35-18 | 86-48 | 0 | 6-00 | 6-90 | 0-00 | 6-70 |
| 1973 | 8 | 16 | 3 | 58 | 23-05 | 101-11 | 0 | 5-40 | 6-40 | 0-00 | 0-00 |
| 1974 | 5 | 10 | 19 | 25 | 28-24 | 104-01 | 11 | 6-20 | 6-80 | 0-00 | 6-70 |
| 1974 | 8 | 11 | 1 | 13 | 39-45 | 73-83 | 9 | 6-40 | 7-30 | 0-00 | 6-90 |
| 1975 | 1 | 19 | 8 | 2 | 32-45 | 78-43 | 33 | 6-20 | 6-80 | 0-00 | 6-80 |
| 1975 | 10 | 3 | 5 | 14 | 30-25 | 66-31 | 11 | 5-80 | 6-70 | 0-00 | 6-70 |
| *1976 | 4 | 8 | 2 | 40 | 40-31 | 63-77 | 33 | 6-50 | 7-00 | 0-00 | 7-10 |
| 1976 | 5 | 29 | 12 | 23 | 24-57 | 98-95 | 8 | 6-10 | 6-90 | 0-00 | 6-90 |
| 1976 | 5 | 29 | 14 | 0 | 24-53 | 98-71 | 10 | 6-00 | 7-00 | 0-00 | 6-90 |
| 1976 | 8 | 12 | 23 | 26 | 26-68 | 97-07 | 27 | 6-40 | 6-10 | 0-00 | 0-00 |
| 1976 | 8 | 16 | 14 | 6 | 32-75 | 104-15 | 16 | 6-10 | 6-90 | 0-00 | 0-00 |
| 1976 | 11 | 6 | 18 | 4 | 27-60 | 101-05 | 33 | 5-80 | 6-50 | 0-00 | 6-50 |
| *1977 | 11 | 18 | 5 | 20 | 32-69 | 88-38 | 33 | 5-70 | 6-50 | 0-00 | 6-30 |
| 1978 | 3 | 24 | 21 | 5 | 42-83 | 78-60 | 33 | 6-20 | 7-10 | 0-00 | 0-00 |
| 1978 | 11 | 1 | 19 | 48 | 39-34 | 72-61 | 40 | 6-20 | 6-80 | 0-00 | 6-80 |
| 1980 | 7 | 29 | 14 | 58 | 29-59 | 81-09 | 18 | 6-10 | 6-50 | 0-00 | 6-60 |
| 1981 | 1 | 23 | 21 | 13 | 30-92 | 101-09 | 33 | 5-70 | 6-80 | 0-00 | 6-60 |
| 1982 | 1 | 23 | 17 | 37 | 31-69 | 82-24 | 33 | 6-00 | 6-50 | 0-00 | 6-20 |
| 1982 | 12 | 16 | 0 | 40 | 36-14 | 69-01 | 36 | 6-20 | 6-60 | 0-00 | 6-90 |
| 1983 | 4 | 18 | 10 | 58 | 27-79 | 62-05 | 64 | 6-50 | 0-00 | 0-00 | 6-50 |
| 1983 | 6 | 24 | 7 | 18 | 21-72 | 103-28 | 18 | 6-10 | 6-60 | 0-00 | 6-40 |
| *1984 | 3 | 19 | 20 | 28 | 40-32 | 63-35 | 15 | 6-50 | 7-00 | 0-00 | 7-10 |
| 1985 | 7 | 29 | 7 | 54 | 36-19 | 70-89 | 99 | 6-60 | 0-00 | 0-00 | 0-00 |
| 1985 | 8 | 23 | 12 | 41 | 39-43 | 75-22 | 7 | 6-40 | 7-30 | 0-00 | 7-50 |
| *1986 | 8 | 20 | 21 | 23 | 34-57 | 91-63 | 33 | 5-40 | 6-50 | 0-00 | 0-00 |
| 1988 | 8 | 6 | 0 | 36 | 25-15 | 95-13 | 91 | 6-80 | 7-20 | 7-30 | 0-00 |
| 1988 | 8 | 20 | 23 | 9 | 26-75 | 86-62 | 57 | 6-40 | 6-60 | 6-80 | 6-50 |
| 1988 | 11 | 6 | 13 | 3 | 22-79 | 99-61 | 18 | 6-10 | 7-30 | 7-00 | 6-90 |
| 1990 | 3 | 25 | 14 | 17 | 37-01 | 72-95 | 32 | 6-00 | 6-70 | 0-00 | 0-00 |
| *1990 | 4 | 26 | 9 | 37 | 35-96 | 100-22 | 33 | 6-60 | 6-90 | 0-00 | 6-70 |

of mainshocks in an area is constant. For one set of functions, designated N_1 , L_1 and Z_1 , the constant is $n_1 \text{ yr}^{-1}$ and it is $n_2 \text{ yr}^{-1}$ for another set N_2 , L_2 and Z_2 . These constants n_1 and n_2 are chosen as 1 and 2 in the whole of the study, if they are not mentioned specifically. For all six functions the duration of a time interval $S=6$ years is chosen. The upper magnitude threshold \bar{M} is adjusted to M_0 . For functions Z , $\bar{M} = M_0 - 0.2$ and for B the threshold were $\bar{M} = M_0 - 2.0$ and $\bar{M} = M_0 - 0.1$ (Bhatia *et al* 1989). A more detailed description is given in Keilis-Borok and Kosobokov (1986, 1990).

The seismic region under consideration is scanned by overlapping squares of radius

300 km to calculate the above functions. Although the diameter (L , in degrees) depends on M_0 , as the empirical relation suggests, $L = \exp(M_0 - 5.6) + 1$ (Keilis-Borok and Kosobokov 1990), we have used a fixed radius to overcome the problem of low annual seismic rate in the region. Now in a square the earthquake flow at a time t is represented by a vector of seven functions evaluated from the preceding time. The problem is to deduce whether t starts the TIP for a strong earthquake. The duration τ of a TIP was chosen to be 5 years (Keilis-Borok and Kosobokov 1986). To declare a TIP at time t we require that over the preceding 3 years (including t) –

1. Each group $\{N_1, N_2\}$, $\{L_1, L_2\}$, $\{Z_1, Z_2\}$ and $\{B\}$ contains functions that have extremely large values.
2. At least six of the seven functions have extremely large values.

Here ‘extremely large’ means values in the upper $Q\%$ quantile ($Q = 10$ for all functions but B and $Q = 25$ for B). The vector is computed at discrete times $\{T_j\}$ using a 6 month step. The interval $(T_j, T_{j+\tau})$ is diagnosed as a TIP if the condition above occurs at T_{j-1} and T_j .

While applying M8 algorithm in this territory we made the following modifications:

- (i) The magnitude threshold is taken to be 6.4, as there are very few earthquakes of $M \geq 7$ and none of $M \geq 8$ after 1964.
- (ii) We do not terminate the TIP after the occurrence of strong shock. This was done because in some regions (like in Indo-Burma) the strong earthquakes are not well separated in time as well as in space.
- (iii) The annual seismicity rate is high in Hindukush-Pamir and Indo-Burma regions but very low in the Himalayan region. So we could not use the standard parameters (10 and 20) in M8 algorithm. Even in high seismicity regions like Hindukush, the level is lower than the standard. But we tried to use the maximum level in a square, and uniform in a region.

3. Results and Discussion

To make the discussion simpler, we divide the seismic belts in four groups (figure 2):

1. Seismic belts in India-Burma and China region.
2. Himalayan seismic belt.
3. Seismic belts associated with the Hindukush-Pamir mountains.
4. Seismic belt in Pakistan region.

Each region with the results is discussed in detail in the following passages.

3.1 Indo-Burma and China region

Most of the seismicity is associated with the N–S trending Arakan-Yoma and almost NW–SE trending Red river and Kang Ting faults. The Arakan Yoma fault system is characterized by a deep and high level of seismicity as compared to the shallow and low level seismicity along the Red river and Kang Ting faults in the east.

a) In the region of Red river and Kang Ting faults, ten earthquakes ($M \geq 6.4$) since 1971 have occurred. Using the catalogue from 1964 to August 1990, we could predict

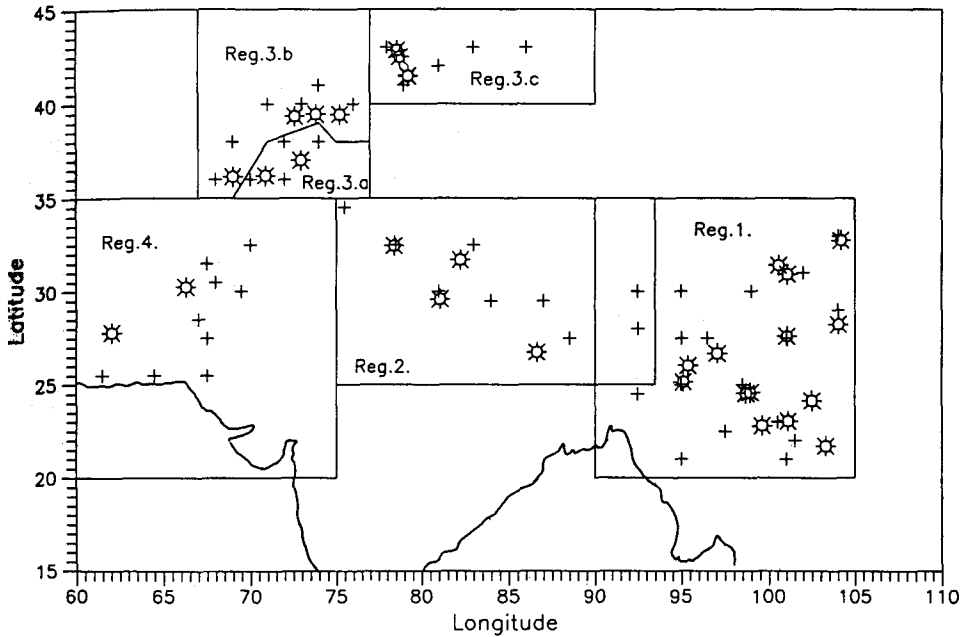


Figure 2. The stars represent the location of the earthquakes ($M \geq 6.4$) occurred in the period from 1970 to 1990) and the centers of the squares are shown by plus signs. The rectangles and the polygons show different regions discussed in the text.

all the earthquakes ($M \geq 6.4$). All these earthquakes were preceded by TIPs (figure 3a) and in all the cases there is a good clustering of foreshocks in time as well as in space, before their occurrence. The earthquake of 70/1/4 could not be predicted as it occurred in the early dates of the catalogue. The TIPs occupy approximately 32% time of the total duration (1970 to August 1990) of analysis. So on an average 3.2% (about 8 months) of the time is used in the TIP for one earthquake.

b). Observing the high annual rate of seismicity along Arakan Yoma arc, we analyzed this region separately by choosing n_1 and n_2 as 3 and 6. Both the earthquakes ($M \geq 6.4$) in the region were predicted (figure 3b). About 17% time of the total, is occupied by the TIPs or in other words 8.5% time (approximately 21 months) per earthquake is used in a TIP. But if we use the catalogue after 1975 the earthquake of 8 August 1988 shows a much bigger area of activation, and it also gives the current alarm in the region which starts on 89/2/4 until 95/2/5 (figure 3c). TIPs occupy 18% time (about 20.7 months) of the analysis (which is from 1981 to August 1990 in this particular case).

Again one earthquake on 70/7/29 did not show the TIP as it occurred at an early date.

3.2 Himalayan region

The seismicity, almost confined between the Main Frontal Thrust and the Indus-Tsangpo suture zone (figure 1), is very shallow and diffused in nature. There hardly exists any good clustering of foreshocks of the four earthquakes ($M \geq 6.4$) in the

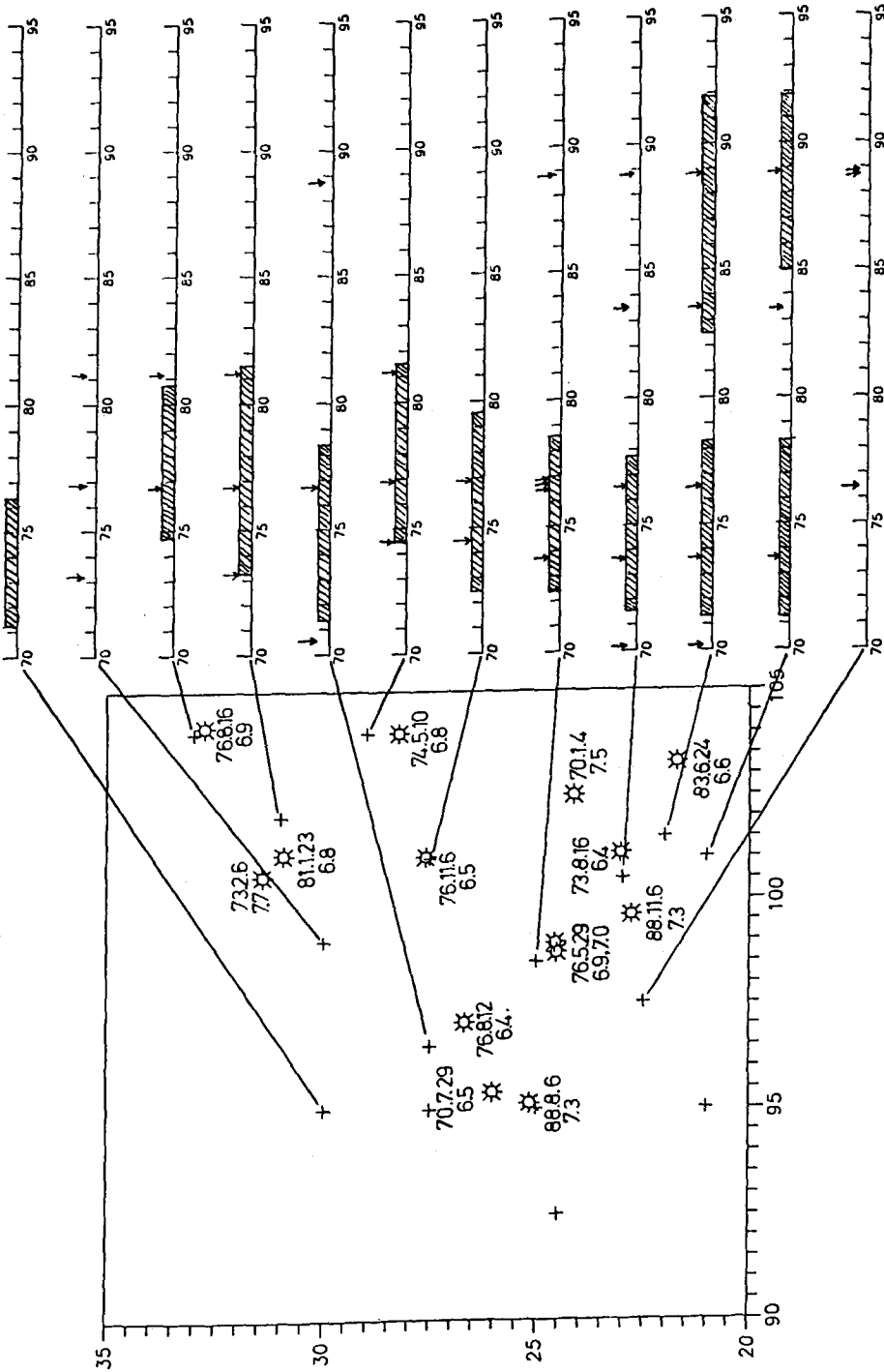


Figure 3a. Indo-Burma and China region (Reg. 1 of figure 2). On the left the distribution of earthquakes and centers of the squares are shown. The numbers represent the dates and the magnitudes of the earthquake. On the right side the TIPs are plotted as the hatched horizontal bars and the arrows indicate the presence of the earthquake in that corresponding square. The numbers indicate the year.

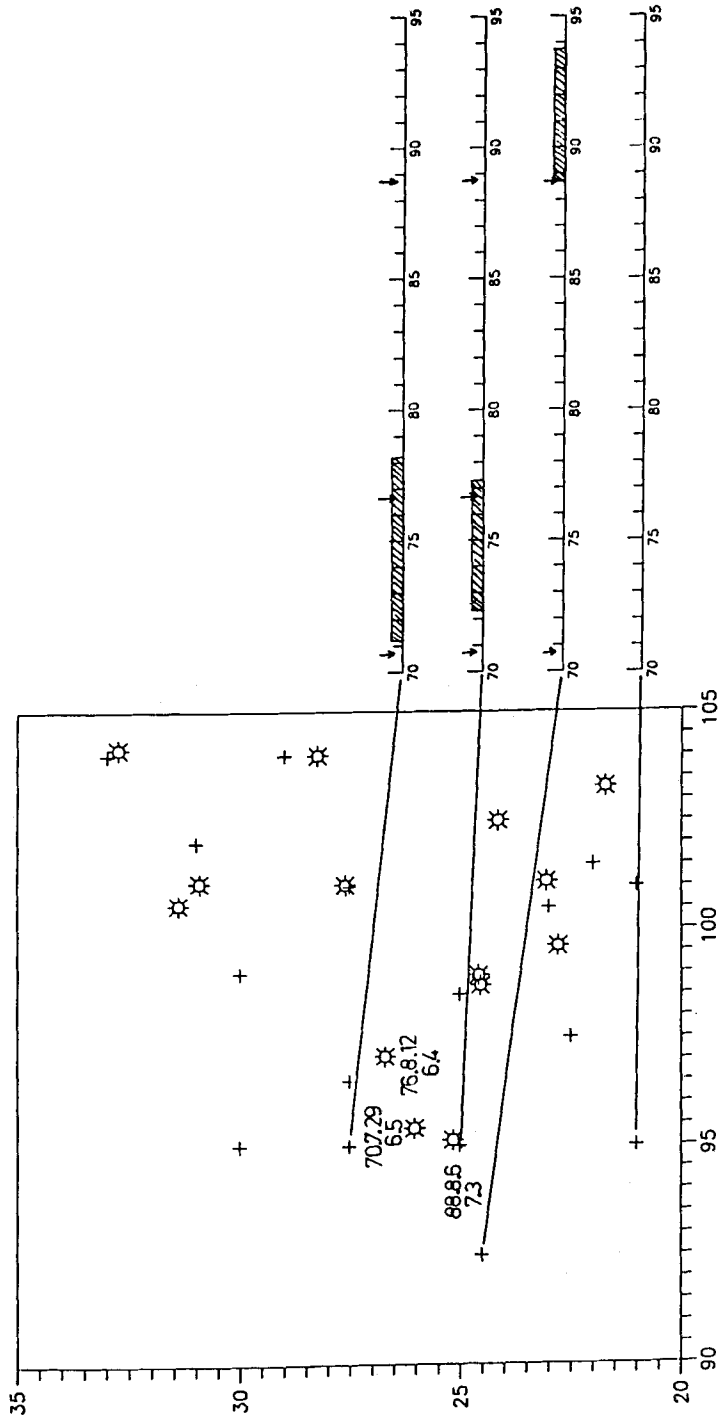


Figure 3b. Arakan-Yoma arc (Reg. 1 of figure 2). Rest of the notations are the same as in figure 3a.

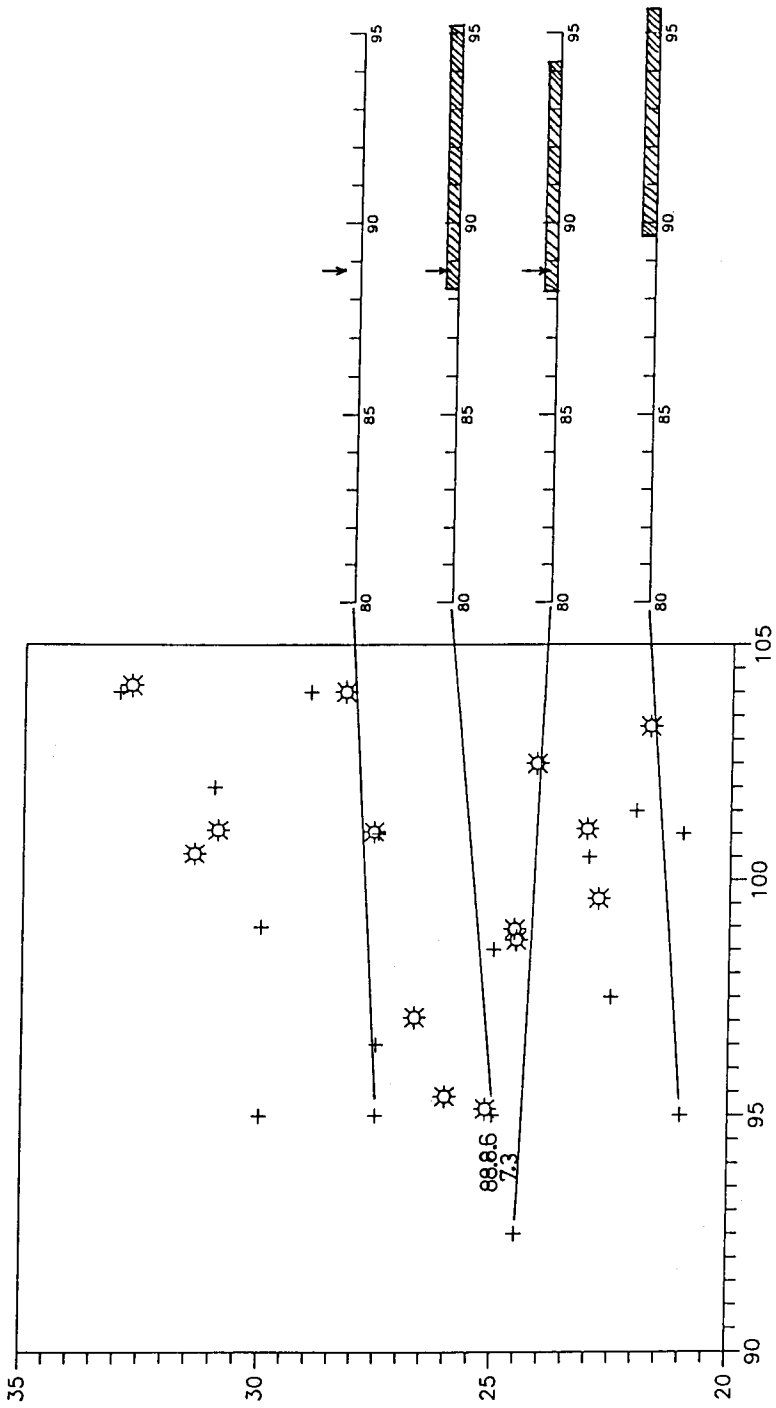


Figure 3c. Area and notations are the same as in figure 3a.

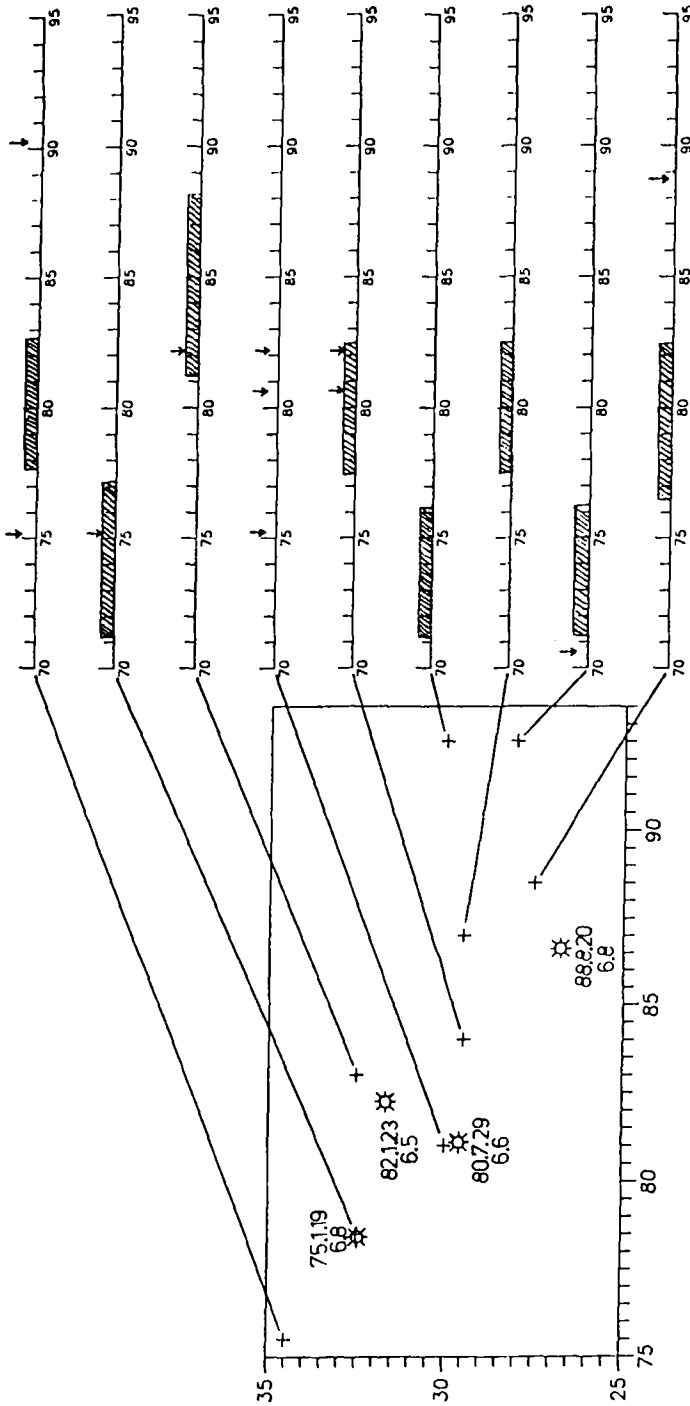


Figure 4. Central Himalaya (Reg. 2 of figure 2). Notations are the same as in figure 3a.

region. Out of these four quakes three were predicted (figure 4), the earthquake that occurred on 20 August 1988 did not show any good precursor. This earthquake had a large number of aftershocks (about 50) but their magnitudes are unknown. The time volume occupied by the TIPs is 23.7% or about 6% time (14.8 months) is used in a TIP for one earthquake.

3.3 Hindukush-Pamir region

In this region of interest there are two distinct belts of seismicity, the southern belt corresponds to a deep and shallow level of seismicity whereas the northern belt is characterized by the shallow events only. There exists a remarkable gap in seismicity in between these two E-W trending belts (figure 5). When these two belts were studied together, the ratio of success to failure to predict the earthquakes was very low but when they were analyzed independently the ratio increases, which means that the data of one belt contaminate the data of the other one. The annual seismicity rate is quite high, hence n_1 and n_2 are chosen to be 5 and 10 respectively.

a) Let us first focus on the southern belt. The two earthquakes ($M \geq 6.4$) that occurred there, shows their predictable nature by giving the TIP (figure 6a). Besides this, one current alarm in the region is diagnosed, starting from 1987 until 1995. TIPs occupy approximately 11% time of the period of analysis or 5.5% (13.5 months) in a TIP for one earthquake.

b) The northern belt is occupied by the four earthquakes ($M \geq 6.4$). All of them are preceded by the TIPs (figure 6b). 39% of the total time is occupied by the TIPs. Hence on an average about 10% time (2 years) is occupied by a TIP for one earthquake.

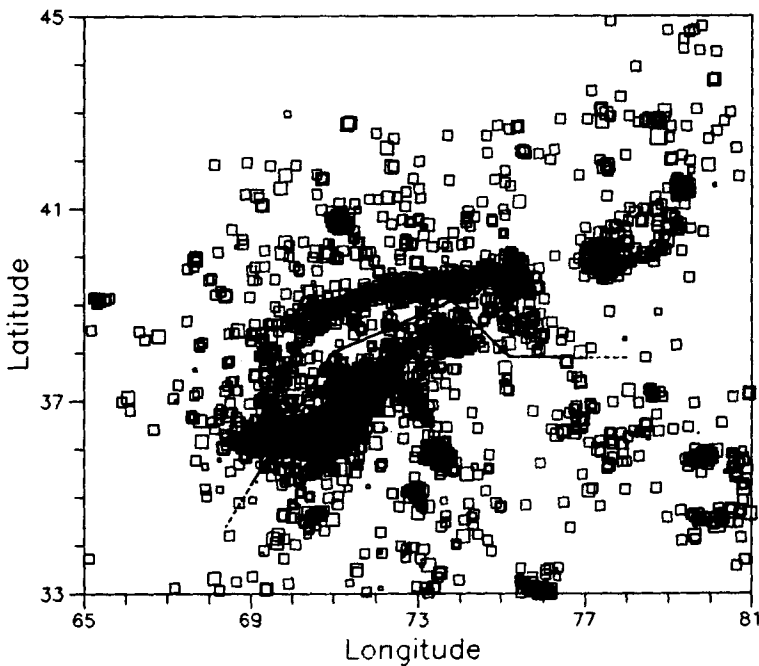


Figure 5. The two distinct belts of seismicity in the Hindukush-Pamir region are shown in this figure.

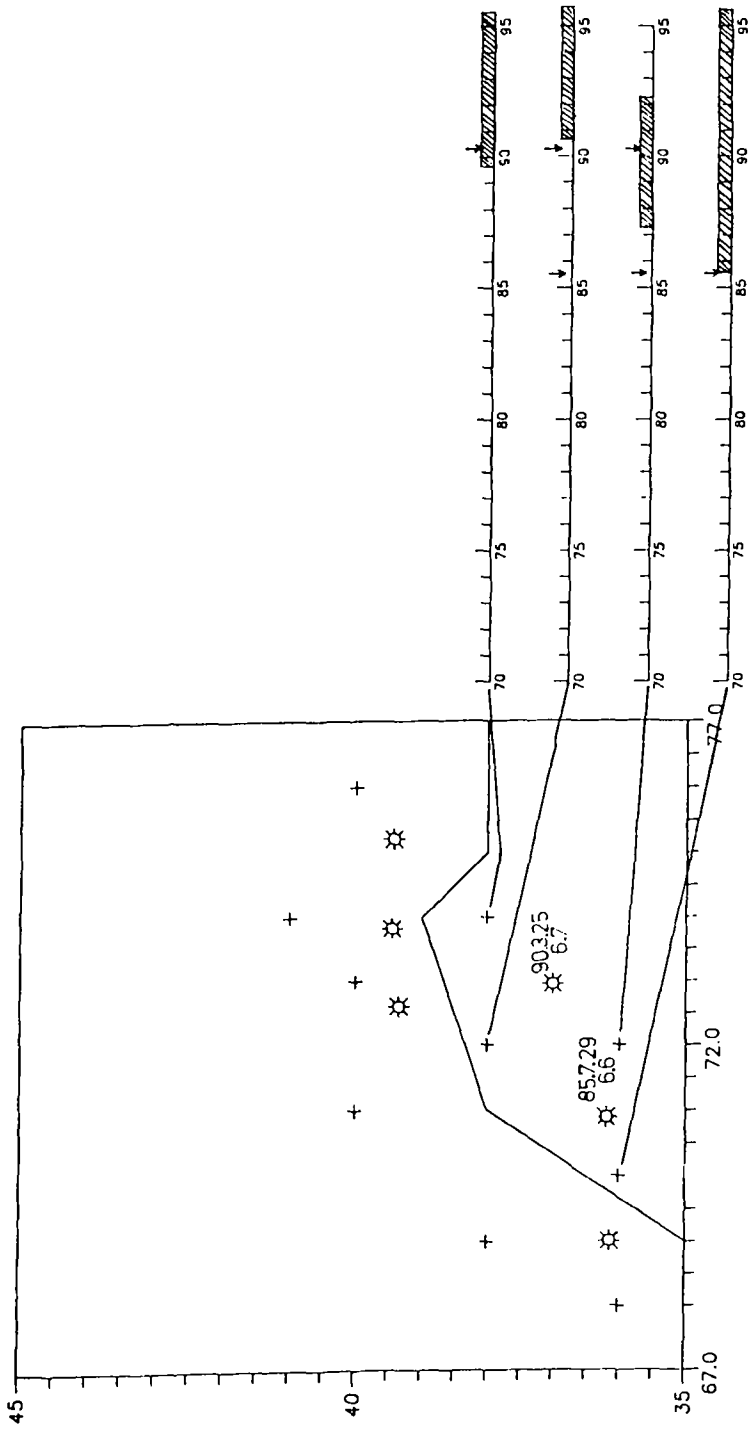


Figure 6a. Southern belt (see figure 5) of the Hindukush-Pamir region (Reg. 3a of figure 2). Notations are the same as in figure 3a.

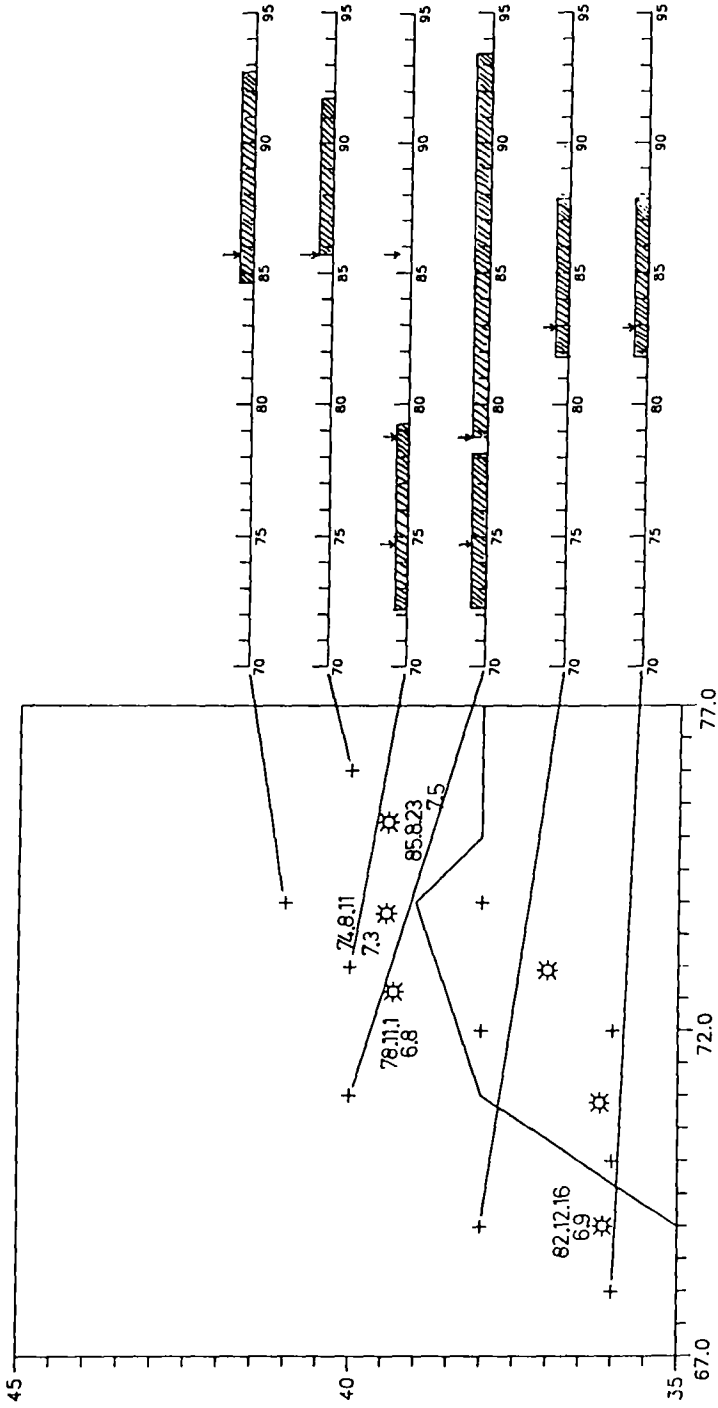


Figure 6b. Northern belt (see figure 5) of the Hindukush-Pamir region (Reg. 3b of figure 2). Notations are the same as in figure 3a.

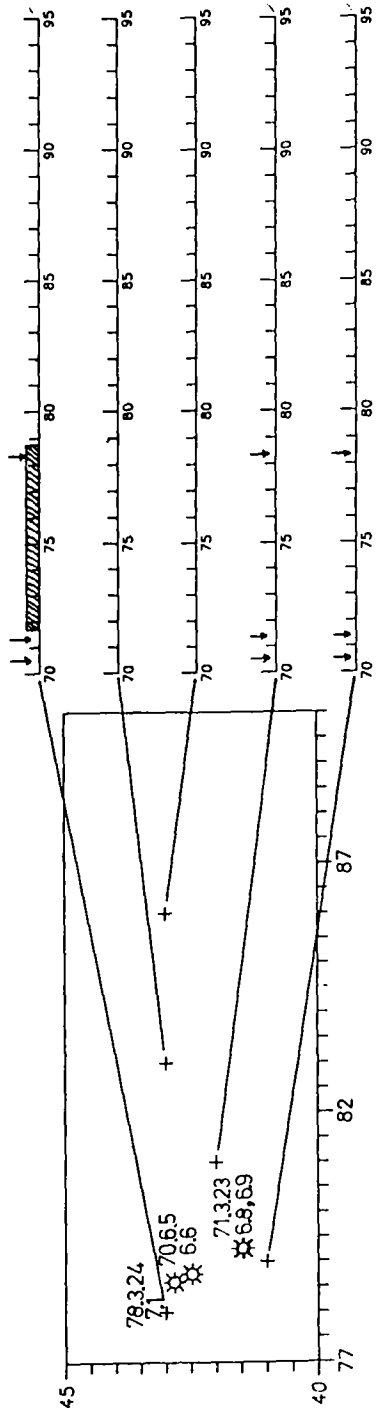


Figure 6c. The eastern part of the Hindukush-Pamir region (Reg. 3c of figure 2). Notations are the same as in figure 3a.

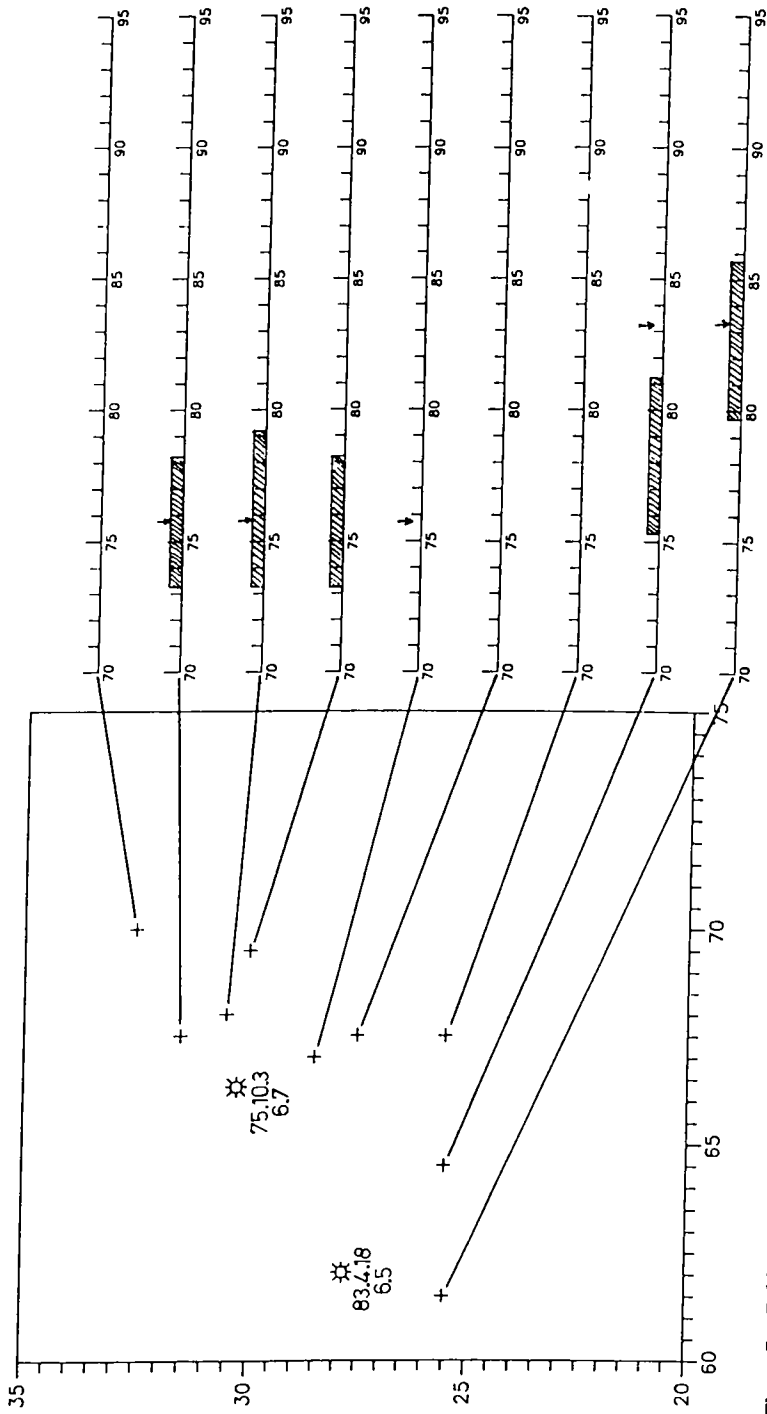


Figure 7. Pakistan region (Reg. 4 of figure 2). Notations are the same as in figure 3a.

c) In the eastern extension of these belts, where n_1 and n_2 are chosen as 3 and 6, the TIP for the earthquake of 24 March 1978, was diagnosed whereas the earthquakes of 5 June 1970 and 3 March 1971 do not show the precursor (figure 6c). In fact this is expected because the dates of these earthquakes fall in the beginning of the catalogue. The percentage of the total time occupied by the TIP is 6.8 (16.8 months).

3.4 Pakistan region

The seismic belt is mainly confined along the Chaman fault and the fold belts (figure 1). There are two earthquakes ($M \geq 6.4$) in the region and one of them broke the Chaman fault on 3 October 1975. The earthquake of 18 April 1983 occurred in the NW of the seismic belt. The TIPs for both the earthquakes are diagnosed (figure 7). The results are stable with n_1 and n_2 as 2 and 4. 15% of the total time is occupied by the TIPs. Or in other words, on an average 7.5% (18.5 months) time is occupied by a TIP for one earthquake.

4. Conclusions

The results show a very high success rate as all the earthquakes ($M \geq 6.4$) after 1971 were predicted except the August 1988 earthquake in the Central Himalaya. Two current alarms, one in Hindukush-Pamir belt and the other in the Arakan Yoma belt were diagnosed.

In all the regions except the Himalaya most of the earthquakes (specially in region 1) show the clustering of foreshocks, which may be as far as 200 km away in space and 3 years ahead in time (figures 3, 4, 6 and 7) from the main major shock.

This analysis strongly supports the predictable nature of the M8 algorithm and recommends its future application on the updated catalogue to increase the possibility of short term prediction.

Acknowledgements

This research was carried out within the framework of the pilot activities of the International Institute for Earth, Environmental and Marine Sciences and Technologies, Trieste, Italy. We are very thankful to Prof G F Panza and Dr I Rotwain for stimulating discussions. One of the authors (VKG) is thankful to Profs V K Gaur and R Chander for their encouragement and continuous interest in the problem. He is grateful to the Chairman of UGC, Engg. & Tech. (India) and the Head of the Earth Sciences Department, Roorkee, for their permission to visit IIEM. We extend our thanks to the directors of IIEM and ICTP for financial support. The suggestion and criticisms of the referees have considerably improved the presentation.

References

- Bhatia S C, Chalam S V, Gaur V K, Keilis-Borok V I and Kosobokov V G 1989 On intermediate-term prediction of strong earthquakes in the Himalayan arc region using pattern recognition algorithm M8; *Proc. Indian Acad. Sci. (Earth Planet. Sci.)* **98** 111-123

- Gabrielov A M, Dimitrieva O E, Keilis-Borok V I, Kosobokov V G, Kuznetsov I V, Levshina T A, Mizroev K M, Molchan G M, Negamatullaew S Kh, Pisarenko V F, Prozorov A G, Rinehart W, Rotwain I M, Shebalin P N, Shnirman M G and Schreider S Yu 1986 *Algorithms of long-term earthquakes prediction*. International School for research oriented to Earthquake Prediction Algorithms, Software and data handling (Lima, Peru, September, 1986).
- Keilis-Borok V I, Knopoff L, Rotwain I M and Sidorenko T M 1980 Bursts of seismicity as long-term precursors of strong earthquakes; *J. Geophys. Res.* **85** 803–811
- Keilis-Borok V I and Kosobokov V G 1986 Time of increased probability for the great earthquakes of the world; *Comput. Seismol.* **19** 48–58
- Keilis-Borok V I and Kosobokov V G 1990 Premonitory activation of earthquake flow: Algorithm M8; *Phys. Earth Planet. Inter.* **61** 73–83
- Tapponier P and Molnar P 1977 Active faulting and tectonics in China; *J. Geophys. Res.* **82** 2905–2930
- World's hypocenters data file 1900–1991, USGS-NEIC, USA, 1990