Mausam (1979), **30**, 2 & 3, 267-273

550.341.5 : 519.24 (541)

Estimated earthquake Probabilities in the northeast India and Andaman-Nicobar islands

P. SITAPATHI RAO and B. RAMALINGESWARA RAO

Department of Geophysics, Andhra University, Waltair

ABSTRACT. Gumbel's extreme value theory has been applied to estimate the probability of occurrence and return periods of the largest earthquakes in the northeast India and Andaman - Nicobar Islands. The statistical model of Epstein and Lomnitz (1966) is discussed with reference to the Gumbel's extreme value theory. The mean lines of expected extremes based on 46 years data of yearly extreme values of earthquakes in the three regions are plotted separately and the mean return periods of the largest possible earthquakes with their probability of recurrence are estimated. The most probable largest earthquakes in the Richter scale that may occur in an year and in an interval of 50 years are also estimated and reported.

1. Introduction

Among the Stochastic models used in the estimation of earthquake risk, the extreme value methods, originally proposed by Gumbel during the 1930's for the flood analysis, have been applied in recent years to earthquake data for obtaining recurrence period and probability of occurrence of the largest earthquakes. In this, given a function $G(x, t)$ of a random variable x on time scale t and is divisible into equal intervals of time τ, each segment contains among a minimum and a maximum in the interval τ. The maximum value $Y=[(x_{max})]_\tau$ called here the extreme value is an independent earthquake event with the highest magnitude in the interval τ (=one year). In a population of data collected over an appreciably long period, an extreme is obtained in each of the interval spiked. Since the true probability of occurrence of these extremes may be estimated with their return periods, a number of authors have applied the theory of extremes and analysed earthquake data in different regions of the world. The first attempt was, however, by Nordquist (1945) on the earthquakes of southern California and also on the largest earthquakes of the world. His observed distribution of the magnitude of the largest earthquakes was found to be in good agreement with the extremal theory, envisaged. Over the times, as the statistical theories and models are in increasing use in all the fields of geophysics, Arnold Court (1952) has brought out a full review of the theory of extremes and its importance to help the civil engineering designs. However, Gumbel (1958) published a complete theory of the extremes and its application.

In the last one decade, extremal value theory was developed and applied by Gayskiy and Katok (1965) in Soviet Union and by Epstein and Lomnitz (1966) in America. Dick (1965) used it for the analysis of the New Zealand earthquakes, Milne and Devenport (1969) for the Canadian earthquakes, Karnik and Hubernova (1968) and Schenkova and Karnik (1970) for the European earthquakes. For the North Circum-Pacific seismic belt, Shakal and Willis (1972) have applied this theory and estimated the earthquake return periods and probabilites as envisaged by Gumbel (1958) in his model. In general, the extreme value methods, their limitations and the earthquake risk involved have been brought forth by Lomnitz (1974) in two articles.

On the earthquake data of the Indian sub-continent, hitherto no application of Gumbel's extreme value theory has been made and Rao and Rao (1978) has reported elsewhere the preliminary analysis and results on the Indian Ocean seismic belts. An extension of this work to northeast India-Andaman-Nicobar Islands forms the subject matter of this paper.

2. Theoretical model

The theoretical model originally proposed by Gumbel (1941) for the analysis of flood data is based on the random variable function $G(x, t)$. It requires n independent observations collected continuously over an appreciably long time which should be amenable for division into N-number of independent sets each having an equal time-length τ. The N sets obtained include N number of extremes as each set contributes invariably an extreme. The parent population must follow, as proposed, a known statistical distribution such as the normal, exponential, chi-square or gamma distribution. Thus, the earthquake data may be modelled into N-sets and the largest magnitude earthquakes may be picked up from each of the

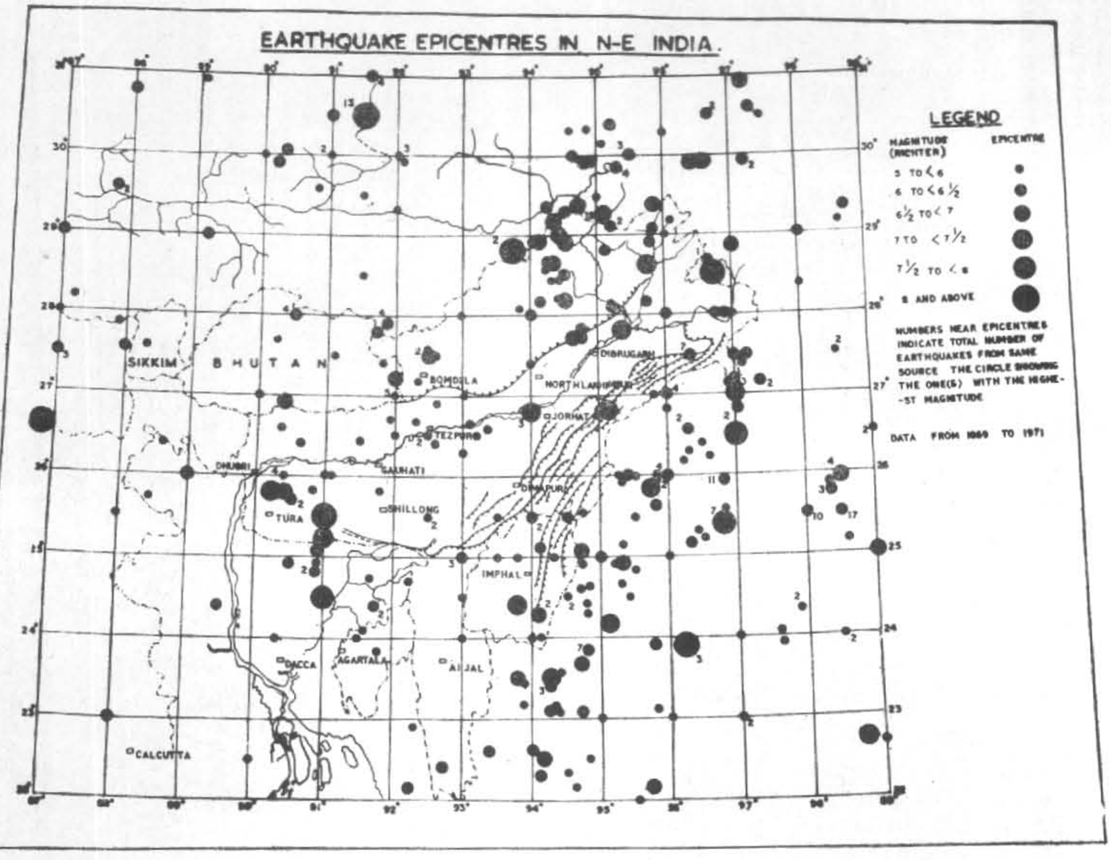

Fig. 1. The epicentral location of earthquakes or M>5; prepared and published by IMD (1974) for Assam and northeast Ind;a

N sets. The earthquakes with largest magnitude, $y_1, y_2, \ldots\ldots y_N$ taken from the N-number of sets are arranged in the increasing order of magnitudes. From the fundamental theorem of the theory of extremes, it is known that as both n and N grow large the cumulative probability where any of these N extremes will be less than any chosen quantity y reaches the double exponential expression. The frequency of each y_i in the ordered set of extremes and the cumulative probability may be represented respectively .

$$G\,(y_i) = \frac{i}{(N+1)} \qquad (1)$$

$$G(y) = Q\,(y) = \exp\,[\,-\exp - \beta\,(y - u)] \qquad (2)$$

where $Q(y)$ is the non-occurrence of an event in a single trial and thus the probability of occurrence $P = 1 - Q = 1 - G(y)$. The expression (2) gives the probability of non-occurrence of an event y in a single trial. The return period of extremes may be consequently obtained for values equal to or exceeding y as

$$T_y = 1/[1 - G\,(y)] \qquad (3)$$

Further, the variation of the probability of y may be obtained by differentiation of the equation (2) and may be written

$$G'(y) = \beta \exp\,[-\,\beta\,(y - u)]\; G\,(y) \qquad (4)$$

Further, the second differential of it gives the maximum density of probability at $y = u$ which means that u is the theoretically the most frequent value (mode) of the set of extremes under consi deration. β is considered theoretically a measure of concentration about the mode, but practically it is obtained by the theory of least squares from the data of the sample or by using two theoretical quantities. :

$$\beta = \frac{\sigma_N}{S_y} \;\; \text{and}\; u = \bar{y} - S_y\,(\bar{R}_N/\sigma_N) \qquad (5)$$

where $\bar{y}$ is the mean value and S_y the standard deviation of the set of extremes, while the reduced mean R_N and the standard deviation σ_N of a theoretical variate which depends on the sample size N. The modal extreme u is always less than the mean value y for the set of extremes. By taking double logarithms of the exponential equation (2), the reduced variate R becomes

$$R = -\ln\,[\,-\ln G(y)] \qquad (6)$$

$$= \beta\,(y - u) \qquad (7)$$

When solved for y, the above equation may be written as :

$$y = u + R/\beta \quad (8)$$

By substitution of the equation (5) for β and u, the equation (8) becomes

$$y = \bar{y} + (S_y/\sigma_N)(R - \bar{R}_N) \quad (9)$$

Which gives a line of expected extreme, if plotted for any set of N extremes.

Without referring to the theoretically calculated values of the variate as shown above, the equation (2) may be changed into by assuming $\ln \alpha = \beta u$; then

$$G(y) = \exp[-\alpha \operatorname{Exp}(-\beta y)] \quad (10)$$

By taking logarithms twice on both sides, equation (10) may be written as :

$$\ln[-\ln G(y)] = \ln \alpha - \beta y \quad (11)$$

The cumulative probability $W_y(k)$ in this case when an extreme value equalling or exceeding y occurs before or along the kth observation, may be written as :

$$W_y(k) = 1 - [G(y)]^k$$
$$= 1 - \exp[-\alpha k \exp(-\beta y)] \quad (12)$$

A similar model but with an assumption that the number of earthquakes in a set (per year) has to follow Poisson distribution was proposed by Epstein and Lomnitz (1966) for the occurrence of large earthquakes. A case of poor Poisson distribution would mean that the region under investigation is inhomogeneous and alternatively a best fit shows the homogeneity of the region. Since Epstein and Lomnitz model requires all the minor and major earthquake events in a region for over a period of atleast a few years to test the *priori* conditions of Poisson's distribution, Gumbel's model is considered in the present paper and its application on the regions of NE-India, Assam and Andaman—Nicobar Islands. For the test of the best fit of the actual extreme values picked up from each set, Gumbel (1958) suggests that it may be obtained as given in the equation (10) by plotting the confidence bands or control curves on a specially prepared extremal probability paper which has a double logarithmic abscissa and a linear ordinate. The return periods on the upper scale and the probabilities on the lower scale for the given extreme magnitudes may be read from the graph. Thus, every extreme value y_i could be described by the confidence level as given by Gumbel (1958) within the frequency limit of 0.15 and 0.85 of $G(y)$ as $(y_i) \pm \sigma(y_i)$ where

$$\sigma(y_i) = -[1/G(y) - 1]^{\frac{1}{2}} / [\beta \sqrt{N} \ln G(y)] \quad (13)$$

Beyond the limit of 0.85, $\sigma(y)$ has the value $1.14078/\beta$ and $0.75409/\beta$ at $G(y_N)$ and $G(y_{N-1})$ respectively. If the observed extremes fall within the band width $2\sigma = 0.45$, they are said to be represented adequately by the probability theory of the extremes. Confidence bands thus constructed for control of the extremes would give the reliability of the test between the theoretical distribution and the true distribution.

3. Application and discussion of results

The model has been worked out to estimate the return periods and probabilities on the regions of northeast India, Assam and Andaman-Nicobar Islands. The shallow focussed earthquake data analysed here covers a period of 46 years starting from 1926 to 1971 and has been taken from the documents of India Meteorological Department (IMD) and from the earthquake catalogues prepared by Tandon and Chatterjee (1968) and later by Tandon and Srivastava (1974). Fig. 1 gives the epicentral location of earthquakes of magnitude greater than 5 as prepared by IMD for Assam and NE India and Fig. 2 the epicentral location of Andaman-Nicobar Islands region. These regions are selected on the basis of their tectonic importance which has been geologically described in the literature by main faults, folds and thrusts with their tributaries. An appraisal of the Indian tectonics including the NE India in recent times was given by Eremenko and Negi (1968). It would be desirable for analysis, had the data covered a much longer period than 46 years, but the earthquake events of the earlier years were sporadically reported (Oldham 1883). Therefore, the earthquake reports earlier to 1926 are not included and the seismic nature of the area is considerably borneout by the 46 year data as revealed in this study. Further the regions are almost exhibiting a homogeneous earthquake process, although geologically and in mechanism they pose complex problems. In the present analysis, the random events of the earthquakes spread over 46 years are divided into equal time intervals of one year. Spiking the interval shorter than one year has made the data discontinuous in the intervals. Gumbel (1958) proposes for the model that the extreme value in a set is independent from the other and the one year time interval ensures the extremes to be independent.

The region, Assam, considered here in particular consists of two seismic zones well known in the literature as lower Assam extending within the coordinates 24.5°N to 26.5°N and 89.5°E to 93.5°E, and upper Assam within the coordinates 27°N to 30°N and 95°E to 98.5°E. Besides thess two blocks, the NE-India as a whole comprising of Assam, Burma and eastern parts of Himalayas is considered as a single block for the analysis

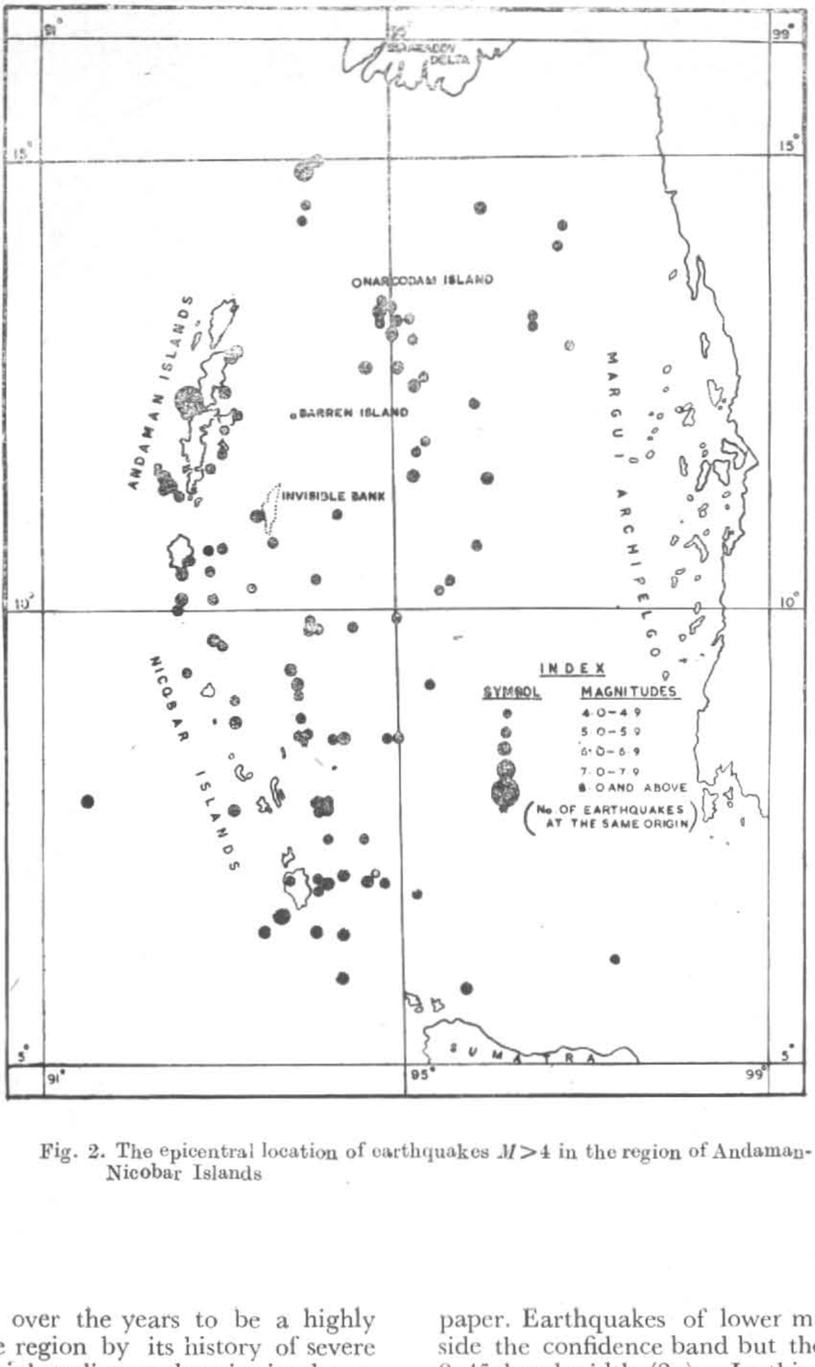

Fig. 2. The epicentral location of earthquakes $M>4$ in the region of Andaman-Nicobar Islands

Assam is known over the years to be a highly earthquake prone region by its history of severe earthquakes. Special studies on the seismic phenomenon of this region have been made by many authors. Montessus De Ballore (1904), Pendse (1948), Tandon (1956), Banerji (1957), Kaila, Gaur and Harinarain (1972), Chaudhury (1973) and Tandon and Srivastava (1974) were among the authors who worked on the seismicity of India with special reference to the Assam region.

Fig. 3(a) gives the extreme value distribution of earthquakes for the Assam region with its confidence bands plotted on the extremal probability paper. Earthquakes of lower magnitude fall outside the confidence band but they are within the 0.45 band-width (2σ). In this region the mean return period of an earthquake of magnitude $M\geqslant8.5$ as obtained from the graph is hundred years. Fig. 3(b) gives a plot of the observed extremes for the NE India and in this case the distribution of extremes fall within the band-width. It may be observed in the figure that the mean return period of a chosen magnitude $M\geqslant8.5$ is found from the upper scale to be 50 years. Similarly Andaman - Nicobar Island area, shown in Fig. 3(c) the mean return period of $M\geqslant8.0$ is about 80 years.

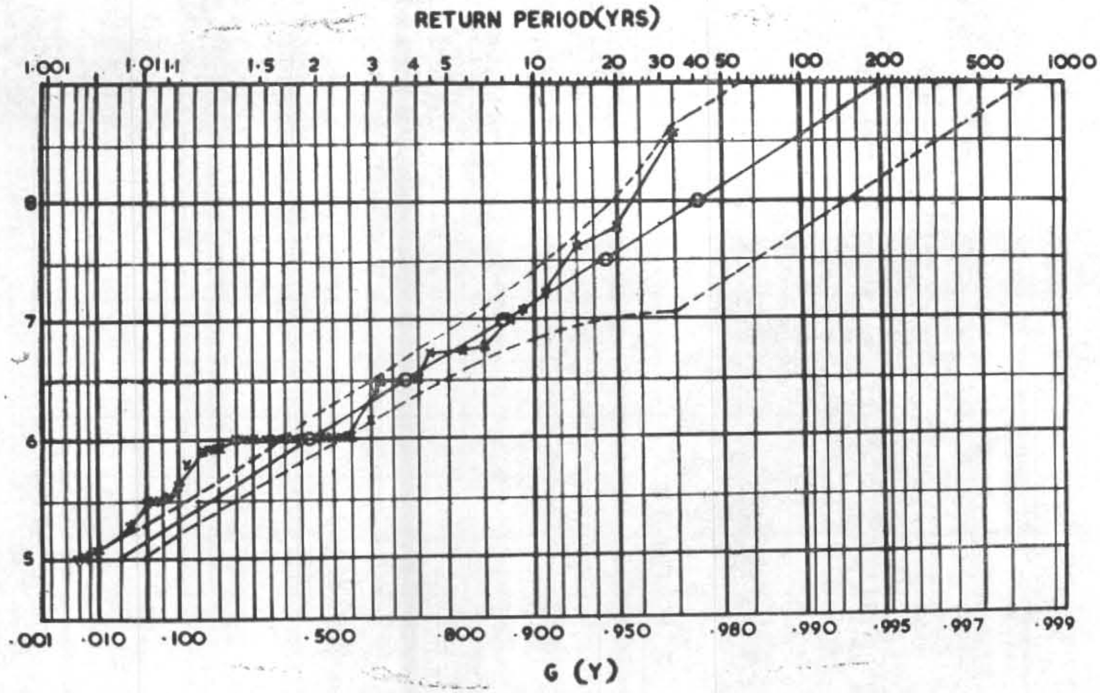

Fig. 3(a). The extreme value distribution of earthquakes with its confidence bands in the Assam region, taken two blocks together

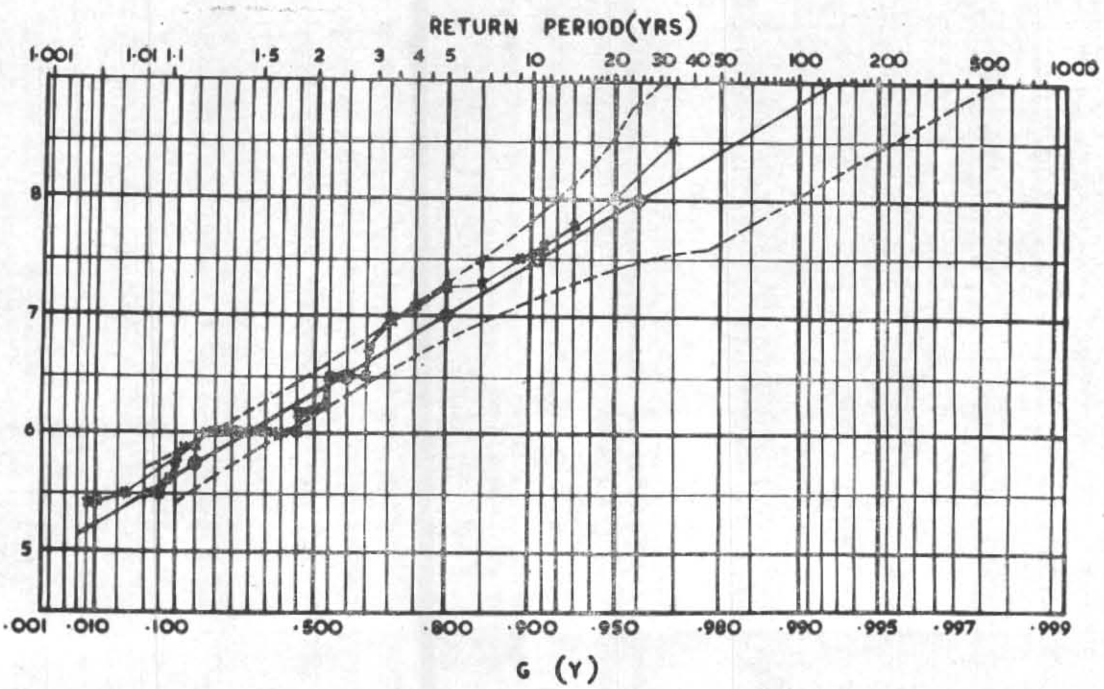

Fig. 3(b). The extreme value distribution of earthquakes with its confidence bands in the region of NE-India

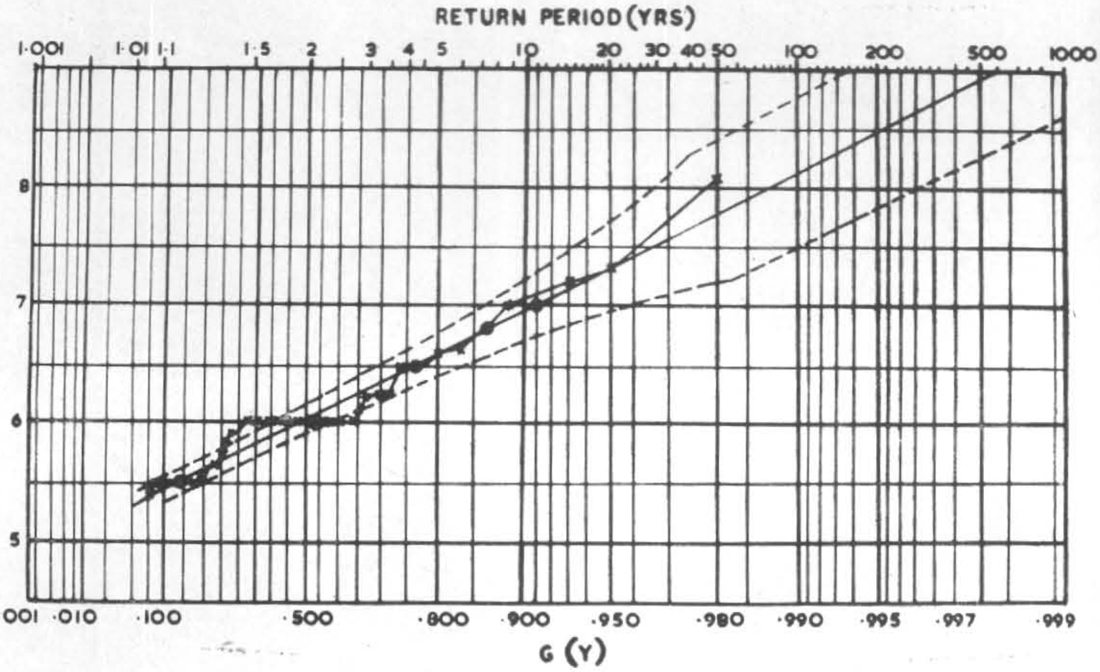

Fig. 3(c). The extreme value distribution of each earthquakes with its confidence bands in the region of Andaman-Nicobar Islands.

TABLE 1

The most probable largest earthquake in one and fifty years for each region with the parameters of least squares

Area	ln α	β	Most probable annual maxima	Most probable 50-year maxima
Assam	9·67	1·66	5·80	8·1
NE-India	9·59	1·58	6·00	8.5
Andaman-Nicobar Islands	11·08	1·93	5·75	7·7

Table 1 summarises the results obtained on the three regions for the least square parameters, ln α and β, and for the most probable largest earthquake in an annual and-fifty years maximal occurrence.

Table 2 gives the mean return periods of an earthquake event in the three regions for the magnitude $M \geqslant 8.0$ and $M \geqslant 8.5$. However, the confidence bands, are relatively wide at higher magnitudes and the return period estimation is thus qualified by the width of the confidence band.

TABLE 2

Mean return periods, percentage probability and the last recorded earthquake events with their magnitudes

S. No.	Magnitude (M) in Richter scale	Mean return period T_y in years	Percentage probability (%)	Year of recodred evidence of last occurrence of the event	Region
1	$\geqslant$8·5	100	37	1897	Assam
2	$\geqslant$8·5	50	43	1950	NE-India
3	$\geqslant$8·0	80	45	1941	Andaman Islands
4	$\geqslant$8.0	25	84	1951	NE India
5	$\geqslant$8·0	42	68	Slightly lower magnitude than 8 occurred in 1947	Assam

The probability calculated from equation (12) for 100 years of return period of an earthquake of magnitude $M \geqslant 8.5$ occurred in the Assam-Shillong plateau is less when compared for a lower magnitude earthquake say about $M \geqslant 8.0$ with its mean return period of 50 years. However, the mean return periods of earthquakes of magnitudes $M \geqslant 8.5$ in NE India and Andaman

Island regions respectively gave higher probability compared to Assam region. Reported in the 4th column of Table 2 are the percentage probability of occurrence of the earthquakes for the highest magnitudes in the three regions.

4. Summary and conclusions

Gumbel's model based on the extreme value theory has been applied experimentally to the earthquake prone regions of Assam, NE India and Andaman island areas of the Bay of Bengal. 46 years of earthquake data which is fairly sufficient for the fulfilment of the envisaged statistical model has been used for the analysis. Assam region including NE India has given below 50% chance of recurrence of an earthquake with the maximum magnitude. At the same time, an earthquake of magnitude about 8, has its probability in the regions exceeding 50% chance with shorter return periods. Andaman Island area also gives probabilities below 50% with its high magnitude earthquake of 8 having a mean return period of 80 years. However, in all the three regions for lack of precise data on the exact measure of magnitude particularly for the earthquakes earlier to 1950 the mean return periods and the probabilities are thus subjected to certain errors, although the mean magnitude curve might help to smooth the errors in the magnitude data. But it may be noted that the slope of the mean magnitude curve which is mainly responsible for the conclusions drawn for the return periods and probabilities is, as known, qualified by the quality of data. It may be said here that the nature of the slope by having precise data particularly for the earlier periods might contribute 1° betterment.

Although low probabilities and large mean return periods for the highest magnitude earthquakes might be an attractive proposition in the civil engineering construction, the conclusions drawn from the studies here in this paper are to be understood within the limitations discussed. Further, as presented in Table 1, the most probable annual maxima of the earthquake magnitudes in the three regions are below 6 on Richter scale and this may warrant certain considerations for the civil engineering constructions.

Acknowledgements

The authors are grateful to Prof. V. Bhaskara Rao, Head of the Department of Geophysics, Andhra University, Waltair for his keen interest and continuous encouragement throughout the progress of the work. The authors are also thankful to the India Meteorological Department and for the help of Drs. A. N. Tandon, H. M. Chaudhury and H. N. Srivastava in providing the data used in this paper. Thanks are also to the Andhra University for awarding the junior author a Research Fellowship to pursue these studies initially.

References

Banerji, S. K., 1957, *Indian Assoc. Cult. Sci.* publ.

Chaudhury, H. M., 1973, Seminar Geodyn. Himalayan Region, N.G.R.I., Hyderabad, pp. 59-71.

Court, A., 1952, *Advance Geophysics*, I, 45-85.

Dick, I. D, 1965, *Proc. World Conf. Earthq. Engng. 3rd*, New Zealand, I, 45-53.

Epstein, B. and Lomnitz., C., 1966, *Nature*, **211**, 5052; 954-956.

Eremenko, N. A. and Negi., B. S., 1968, Oil & Nat. Gas Comm. India : 1-15.

Gayskiy, V. N. and Katok, A. P., 1965, Sbornik Dinamika Zemnoy Kory ANSSR., Nanka, Moskva.

Gumbel, E. J., 1935, *Ann. Inst. Henri Poincare*, **51** : 115-158.

Gumbel, E. J., 1941, *Ann. Math. Stat.* **12** : 163-190.

Gumbel, E. J., 1958, Columbia Univ. Press, New York.

Kaila, K. A., Gaur V. K. and Harinarain, 1972, *Bull, seismol. Soc. Am.*, **62** : 1119-1132.

Karnik, V. and Hubernova, Z. 1968, *Pur. appl. Geophys.*, **70** : 61-73.

Lomnitz, C., 1974, Elsevier Sci. Publ. Co., New York.

Milne, W. G. and Davenport, A. G., 1969, *Bull. seismol. Soc. Am.*, **59** : 729-754.

Montessus De Ballore, 1904, *Mem. geol. Surv. India*, 35.3.

Nordquist, J. M., 1945, *Trans. Am. Geophy. Un.*, 26-I : 29-31.

Oldham, R. D., 1883, *Mem. geol. Surv. India.* **10** :1 -129

Pendse, C. G., 1948, *India met. Dep., Sci. Notes*, **10** : 129-221.

Rao, P. S. and Rao, B. R., 1978, *Proc. Ass. Expl. Geophys*, 4 (To be published).

Schenkova, Z. and Karnik, V., 1970, *Pur. appl. Geophys.*, **33** : 181-228.

Shakal, A. F. and Wills, D. E., 1972, *Bull. seismol. Soc. Am.*, **62**: 1397-1410.

Tandon, A. N., 1956, *Inidian J. Met. Geophys.*, **7** : 93-95.

Tandon, A. N. and Chaterjee, S. N., 1968, *Indian J. Met. Geophys.*, **19** : 273-280.

Tandon, A. N. and Srivastava, S. N., 1974, *Earthquake Engg. Jai Krishna Vol*: 1-49.