Pure Appl. Geophys. 172 (2015), 2533–2544

DOI 10.1007/s00024-014-0776-0

Pure and Applied Geophysics

CrossMark

Distribution of Earthquake Interevent Times in Northeast India and Adjoining Regions

SUMANTA PASARI[1] and ONKAR DIKSHIT[1]

Abstract—This study analyzes earthquake interoccurrence times of northeast India and its vicinity from eleven probability distributions, namely exponential, Frechet, gamma, generalized exponential, inverse Gaussian, Levy, lognormal, Maxwell, Pareto, Rayleigh, and Weibull distributions. Parameters of these distributions are estimated from the method of maximum likelihood estimation, and their respective asymptotic variances as well as confidence bounds are calculated using Fisher information matrices. Three model selection criteria namely the Chi-square criterion, the maximum likelihood criterion, and the Kolmogorov–Smirnov minimum distance criterion, are used to compare model suitability for the present earthquake catalog (YADAV *et al.* in Pure Appl Geophys 167:1331–1342, 2010). It is observed that gamma, generalized exponential, and Weibull distributions provide the best fitting, while exponential, Frechet, inverse Gaussian, and lognormal distributions provide intermediate fitting, and the rest, namely Levy, Maxwell Pareto, and Rayleigh distributions fit poorly to the present data. The conditional probabilities for a future earthquake and related conditional probability curves are presented towards the end of this article.

Key words: Northeast India, Probability distributions, Fisher information matrix, Earthquake recurrence.

1. *Introduction*

The northeast of India and its adjoining regions (20°–32°N and 87°–100°E) have been one of the most seismically active areas over historical time. Since 1846, twenty earthquakes of magnitudes greater or equal to 7.0 have occurred in this region. Among these, two great earthquakes, namely the Shillong Plateau earthquake of 12 June 1897 (*Ms* 8.7) and the Independence Day Assam earthquake of 15 August 1950 (*Ms* 8.6) rocked the whole northeastern region causing extensive loss of human life and property in the Indian subcontinent (GUPTA *et al*. 1986; GUPTA and SINGH 1986; BILHAM and ENGLAND 2001).

Statistical properties of time intervals between successive earthquakes in northeast India and its surrounding regions have been the subject of numerous studies in order to provide long-term prediction for the next big earthquakes. A number of scientists, namely PARVEZ and RAM (1997), YADAV *et al.* (2010), PASARI and DIKSHIT (2013) have earlier carried out recurrence interval estimation of the study region. They used four probability distributions, namely exponential, gamma, lognormal, and Weibull (two-parameter and three-parameter) distributions in their analysis. Apart from these probability distributions, the ones that are commonly used in recurrence modeling are the Pareto group of distributions (KAGAN and SCHOENBERG 2001; FERRAES *et al*. 2003; PISARENKO *et al*. 2010), the Gaussian distribution (PAPAZACHOS *et al*. 1987), the inverse Gaussian or Brownian passage time distribution (MATTHEWS *et al*. 2002; KAGAN 2007), the Rayleigh distribution (FERRAES *et al*. 2003; YAZDANI and KOWSARI 2011), the Levy distribution (SOTOLONGO-COSTA *et al*. 2000), the negative binomial distribution (DIONYSIOU and PAPADOPOULOS 1992), the generalized gamma distribution (BAK *et al*. 2002), and the triple exponential distribution (KIJKO and SELLEVOLL 1981). Nevertheless, the most appropriate distribution function to represent earthquake interevent times still remains under debate. As a result, it has now been a common practice to apply all competing models on a given catalog, and analyze recurrence interval from the best fitted model(s).

In a similar manner, an attempt is made in this study to analyze earthquake interoccurrence times of northeast India and its adjoining regions from eleven probability distributions, namely exponential, Frechet

[1] Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India. E-mail: spasari@iitk.ac.in; sumanta.pasari@gmail.com

(inverse Weibull), gamma, generalized (exponentiated) exponential, inverse Gaussian, Levy, lognormal, Maxwell, Pareto, Rayleigh, and Weibull distributions. Parameters of these distributions are estimated from the method of maximum likelihood estimation (MLE), and their respective asymptotic variances as well as confidence bounds are calculated using the concept of the Fisher information matrix (FIM). The performances of these distributions are evaluated from three statistical criteria: the Chi-square criterion, the maximum likelihood criterion and its modification, named as Akaike information criterion (AIC), and the Kolmogorov–Smirnov minimum distance criterion. In addition to the statistical model developments, we calculate conditional probability values (for different elapsed times) for a large earthquake ($M \geq 7.0$) in the study region.

1.1. *Study Area and Earthquake Data File Description*

In the present work, we investigate earthquake inter-occurrence time by analyzing the northeastern India catalog, including events with magnitude above 7.0 (Yadav *et al.* 2010). This region has been tectonically very active due to the collision and ongoing convergence between Indian plate with Tibet in the north and the Burmese landmass towards the east (Nandy 1986, 2001; Bilham and England 2001). There are a number of active thrust faults, namely main frontal thrust, main boundary thrust, main central thrust, Lohit thrust, Misami thrust, and the Bame-Tuting fault (Gsi 2000). On the basis of epicentral distributions of past earthquakes, faulting pattern, ground evidences, and geotectonic features, the northeast region and its vicinity can be divided into five smaller seismotectonic zones: eastern (upper) Himalayan collision zone, Indo-Myanmar subduction zone, Syntaxis zone of Himalayan arc and Bermese arc (the Mishmi massif), the Brahmaputra valley, and the Shillong plateau (Kayal 1996; Nandy 2001; Thingbaijam *et al.* 2008). Besides, this region, according to the seismic zoning map of India (Bis 2002), falls under zones V, IV, and III, with magnitudes exceeding 8, 7, and 6, respectively.

Table 1 provides a list of 20 major earthquake events ($M \geq 7.0$) from northeast India and its adjoining regions covering a period from 1846 to 1995. The geographical epicentral locations of these events are shown in Fig. 1. It is worthwhile to note here that the study region has not experienced any earthquake of magnitude $M \geq 7.0$ since 1995. Therefore, the present catalog (Yadav *et al.* 2010) essentially accounts for all main shocks with magnitude $M \geq 7.0$ for the period 1846–2013.

2. *Probabilistic Modeling of Earthquake Recurrence*

Let T be a positive random variable of the recurrence time with cumulative distribution function $F(t)$, density function $f(t)$, survival function $S(t)$, and hazard function $h(t)$. Further, we assume that τ is the time elapsed since the last event and v is the waiting time. Having τ as known, waiting time v is random. Thus, our concern is to estimate v so that an earthquake appears within $(\tau,\ \tau + v)$, knowing that no earthquake occurred in the last τ $(0 < \tau < t)$ years. Bringing the concept of reliability into the picture and calling the overall structure as an earthquake system, $S(t)$ illustrates the probability that an earthquake will occur later than time t and $h(t)$ defines the instantaneous rate of earthquake occurrence. $S(t)$ and $h(t)$ are defined by $S(t) = 1 - F(t)$ and $h(t) = \frac{f(t)}{1-F(t)}$.

We further introduce a random variable V corresponding to the waiting time v. Noting that V is linearly related to the random variable T (through the elapsed time τ), its distribution function becomes $F(\tau + v)$. Therefore, the conditional probability of an earthquake in time interval $(\tau,\ \tau + v)$, knowing that no earthquake occurred in the last τ years, can be defined as

$$P(V \leq \tau + v | V \geq \tau) = \frac{F(\tau + v) - F(\tau)}{1 - F(\tau)} \quad (v > 0). \tag{1}$$

Eleven different probability distributions are considered in this study. These distributions and their probability density functions are presented in Table 2. The associated model parameters along with their generic role are also highlighted. The genesis of these distributions (except, generalized exponential), their model properties, and interrelations among

Table 1

List of earthquakes that occurred in the study region for magnitude $M \geq 7.0$ (after Yadav *et al.,* 2010*)*

S. No	Date						Location		Focal depth (km)	Magnitude
	Year	Month	Day	Hour	Min	Sec	Latitude (N)	Longitude (E)		
1	1846	12	10				26.00	93.00		7.5
2	1868	6	30				24.50	91.50		7.5
3	1885	1	1				25.40	90.00		7.3
4	1897	6	12	11	5		25.90	91.90	60	8.7
5	1908	12	12				26.50	97.00		7.5
6	1912	5	23	2	24		21.00	97.00	25	7.9
7	1918	7	8	10	22		24.50	91.00	60	7.6
8	1923	9	9	22	3	42.00	25.25	91.00		7.1
9	1931	1	27	20	9		25.60	96.80	60	7.6
10	1943	10	23				21.50	93.50		7.2
11	1946	9	12	15	20		23.50	96.00	60	7.5
12	1947	7	29	13	43		28.50	94.00	60	7.9
13	1950	8	15	14	9		28.50	96.70		8.6
14	1951	11	18	9	35	45.00	30.50	91.00		8.0
15	1954	3	21	23	42		24.40	95.20	180	7.5
16	1961	2	4	8	51		24.90	93.34	141	7.6
17	1976	5	29	14	0	18.50	24.53	98.71	10	7.0
18	1988	8	6	0	36	24.60	25.15	95.13	91	7.2
19	1991	1	5	14	57	11.50	23.61	95.90	20	7.1
20	1995	7	11	21	46	39.78	21.97	99.20	12	7.1

themselves may be found in Johnson *et al.* (1995). For the generalized exponential, the same may be found in Gupta and Kundu (1999, 2007).

Table 2 shows that the usual domains for ten distributions are the whole positive real line, while for Pareto distribution, the domain (α, ∞) is restricted by the completeness parameter α. Further, three distributions, namely gamma, generalized exponential, and Weibull, when $\beta = 1$, coincide with the exponential distribution, meaning these distributions are somewhat generalizations or extensions of classical exponential distribution. In addition, it is observed (Gupta and Kundu 1999) that generalized exponential distribution shares many physical properties (e.g., shapes of density function or hazard function) with gamma and Weibull models and, thus, could be a potential model to represent earthquake interevent times. Besides, Table 2 includes a number of heavy-tailed (tail is thicker than that of exponential model) distributions: Frechet, Levy, lognormal, Pareto, and Weibull $(\beta < 1)$. As a whole, we have tried to combine all possible type of distributions for a well-defined conclusion for the most appropriate model(s) for the present earthquake catalog of northeast India and the adjoining region.

Apart from the probability density function, characterization of hazard function has become very popular in seismic recurrence studies (Davis *et al.* 1989; Sornette and Knopoff 1997; Matthews *et al.* 2002). The various shapes of hazard functions provide salient information of earthquake reliability. More specifically, increasing hazard function implies that chances of an earthquake increase with time (similar to the elastic rebound theory, Reid 1910), whereas decreasing hazard function means the opposite, and the constant hazard function refers to the chances of an earthquake being independent of elapsed time. It is easy to observe that (a) exponential distribution has constant hazard function, (b) gamma, generalized (exponentiated) exponential, and Weibull distributions possess monotone hazard functions, (c) Frechet (inverse Weibull), inverse Gaussian, lognormal, and Weibull distributions offers both monotone and non-monotone hazard shapes, and (d) Pareto distribution has a decreasing hazard function. Different plots of these hazard functions may be available upon request to the authors.

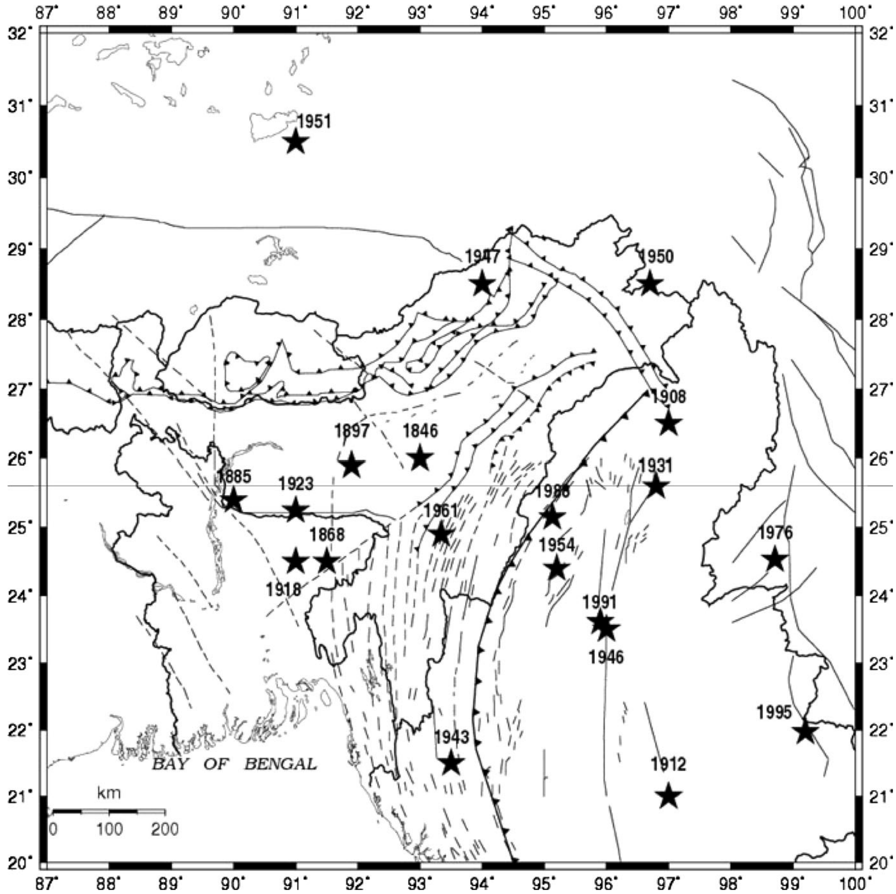

Figure 1
Epicentral locations of earthquakes of magnitude $M \geq 7$ (as listed in Table 1) that occurred in the study region (northeast India and surrounding regions) after 1846 till 2013 (after Yadav *et al.* 2010)

3. Parameter Estimation

Using the interevent times of successive earthquakes (as listed in Table 1), the model parameters of all 11 distributions have been estimated using the MLE method. The detailed procedure of the MLE method may be found in Hogg *et al*. 2005. The estimated model parameter values are shown on Table 4.

Characterizations of the estimated parameters are essential in order to quantify the uncertainty in the estimation process. However, most of the time, the exact distributions of the estimated model parameters are not available. In such cases, the law of large samples is used as a proxy to asymptotically estimate the variance-covariance matrix $\Sigma_{\hat{\theta}}$ and confidence bounds of the estimated parameters $(\hat{\theta})$. In this study, we have exact distribution of the estimated Pareto parameters only (Quandt 1966). The exact variances of the Pareto parameters are given (Quandt 1966) below.

$$\begin{aligned} \sigma^2(\hat{\alpha}) &= \frac{\alpha^2 n\beta}{(n\beta-2)(n\beta-1)^2} \quad \left(n > \frac{2}{\beta}\right) \\ \sigma^2\left(\hat{\beta}\right) &= \frac{\beta^2 n^2}{(n-2)^2(n-3)} \quad (n>3) \end{aligned} \tag{2}$$

For other distributions, we calculate FIM $I\left(\hat{\theta}\right)$ and combine it with the Cramer–Rao lower-bound theorem defined as $\Sigma_{\hat{\theta}} \geq \left[nI\left(\hat{\theta}\right)\right]^{-1}$. A brief discussion on FIM is provided in "Appendix 1". In addition, the Fisher–trace information (FTI), defined as the trace of the FIM, is derived for each

Table 2

Probability distributions and their density functions

Distribution	Density function		Parameters	
	PDF	Domain	Role	Assumption
Exponential	$\frac{1}{\alpha}e^{-\frac{t}{\alpha}}$	$t > 0$	α-scale	$\alpha > 0$
Frechet (inverse Weibull)	$\beta\alpha^{\beta}t^{-\beta-1}e^{-\left(\frac{t}{\alpha}\right)^{-\beta}}$	$t > 0$	α-scale β-shape	$\alpha > 0$ $\beta > 0$
Gamma	$\frac{1}{\Gamma(\beta)}\frac{t^{\beta-1}}{\alpha^{\beta}}e^{-\frac{t}{\alpha}}$	$t > 0$	α-scale β-shape	$\alpha > 0$ $\beta > 0$
Generalized (exponentiated) exponential	$\alpha\beta(1-e^{-\alpha t})^{\beta-1}e^{-\alpha t}$	$t > 0$	$1/\alpha$-scale β-shape	$\alpha > 0$ $\beta > 0$
Inverse Gaussian	$\sqrt{\frac{\beta}{2\pi t^3}}\exp\left[-\frac{\beta(t-\alpha)^2}{2\alpha^2 t}\right]$	$t > 0$	β/α-shape	$\alpha > 0$ $\beta > 0$
Levy	$\sqrt{\frac{\alpha}{2\pi}}\frac{e^{-\frac{\alpha}{2t}}}{t^{3/2}}$	$t > 0$	α-scale	$\alpha > 0$
Lognormal	$\frac{1}{t\beta\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln t-\alpha}{\beta}\right)^2\right]$	$t > 0$	α-log-scale β-shape	$-\infty < \alpha < \infty$ $\beta > 0$
Maxwell	$\sqrt{\frac{2}{\pi}}\frac{t^2}{\alpha^3}\exp\left[-\frac{1}{2}\left(\frac{t}{\alpha}\right)^2\right]$	$t > 0$	α-scale	$\alpha > 0$
Pareto	$\beta\frac{\alpha^{\beta}}{x^{\beta+1}}$	$t > \alpha$	α-scale β-shape	$\alpha > 0$ $\beta > 0$
Rayleigh	$\frac{t}{\alpha^2}\exp\left(-\frac{t^2}{2\alpha^2}\right)$	$t > 0$	α-scale	$\alpha > 0$
Weibull	$\frac{\beta}{\alpha^{\beta}}t^{\beta-1}e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$	$t > 0$	α-scale β-shape	$\alpha > 0$ $\beta > 0$

distribution. The FTI offers an overall measure of the total amount of uncertainty associated with the distribution (GUPTA and KUNDU 2006). The larger the FTI, the better is the approximation.

Now considering the asymptotic normality of the MLE estimated parameters, we derive the asymptotic confidence bounds, or confidence intervals, for each parameter (LAWLESS 1982). The $(1-\delta)\,\%$ two-sided confidence bounds of the parameter θ is obtained as

$$\widehat{\theta} - z_{\delta/2}\sqrt{\mathrm{var}\left(\widehat{\theta}\right)} < \theta < \widehat{\theta} + z_{\delta/2}\sqrt{\mathrm{var}\left(\widehat{\theta}\right)}, \quad (3)$$

$z_{\delta/2}$ is the critical value corresponding to a significance level of $\delta/2$ on the standard normal distribution.

The FIMs and FTIs corresponding to the presently studied distributions are listed in Table 3, whereas the asymptotic standard deviations and confidence bounds are shown in Table 4.

Tables 3 and 4 provide much information related to the distribution properties; for instance, the estimated shape parameters for gamma, generalized exponential, and Weibull models are found to be greater than 1.0, meaning the associated hazard functions of these distributions are monotonically increasing; the FTI for the generalized exponential distribution is the largest, which implies that this distribution is more precise in providing estimated parameter values (which is actually true as can be seen from Table 4). At this point, however, it may also be noted that, for some parameters (e.g., shape parameter of gamma or inverse Gaussian parameters), we see that the numerical figure of the asymptotic standard deviation is quite large. But, it does not necessarily mean that the associated distribution will fit very poorly to the data. Therefore, the emphasis should be on estimating the final propagation in the conditional probability values (discussed later in Sect. 5) with model parameter uncertainties as inputs, rather than deciding directly from the numerical standard deviation values of the estimated parameters.

Table 3

Fisher information matrix (FIM) and Fisher trace information (FTI)

Distribution	Fisher information	
	FIM $(I(\theta))$	FTI
Exponential	$\frac{1}{\alpha^2}$	0.0163
Frechet	$\begin{bmatrix} \frac{\beta^2}{\alpha^2} & \frac{1}{\alpha}(1+\psi(1)) \\ \frac{1}{\alpha}(1+\psi(1)) & \frac{1}{\beta^2}\left(\psi'(1)+\psi^2(2)\right) \end{bmatrix}$	1.5577
Gamma	$\begin{bmatrix} \frac{\beta}{\alpha^2} & \frac{1}{\alpha} \\ \frac{1}{\alpha} & \psi'(\beta) \end{bmatrix}$	2.5315
Generalized exponential	$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ $a_{11} = \frac{1}{\alpha^2}\left[1+\frac{\beta(\beta-1)}{\beta-2}(\psi'(1)-\psi'(\beta-1))+(\psi(\beta-1)-\psi(1))^2\right]$ $-\frac{\beta}{\alpha^2}\left[\psi'(1)-\psi(\beta)+(\psi(\beta)-\psi(1))^2\right]; \beta \neq 2$ $a_{12}=a_{21}=\frac{1}{\alpha}\left[\frac{\beta}{\beta-1}(\psi(\beta)-\psi(1))-(\psi(\beta+1)-\psi(1))\right]; \beta \neq 1$ $a_{22}=\frac{1}{\beta^2}$	86.1161
Inverse Gaussian	$\begin{bmatrix} \frac{1}{2\beta^2} & 0 \\ 0 & \frac{\beta}{\alpha^3} \end{bmatrix}$	0.0247
Levy	$\frac{1}{2\alpha^2}$	0.0360
Lognormal	$\begin{bmatrix} \frac{2}{\beta^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{bmatrix}$	3.8526
Maxwell	$\frac{6}{\alpha^2}$	0.1898
Pareto	$\begin{bmatrix} \frac{\beta}{\alpha^2(\beta+2)} & -\frac{1}{\alpha(\beta+1)} \\ -\frac{1}{\alpha(\beta+1)} & \frac{1}{\beta^2} \end{bmatrix}$	3.7192
Rayleigh	$\frac{4}{\alpha^2}$	0.0844
Weibull	$\begin{bmatrix} \frac{\beta^2}{\alpha^2} & -\frac{1}{\alpha}(1+\psi(1)) \\ -\frac{1}{\alpha}(1+\psi(1)) & \frac{1}{\beta^2}\left(\psi'(1)+\psi^2(2)\right) \end{bmatrix}$	1.0158

4. *Model Selection*

In order to prioritize the competing models, we apply the three model selection criteria, namely the Chi-square criterion, the maximum likelihood criterion and its modifications, known as the AIC, and the Kolmogorov–Smirnov minimum distance criterion. In "Appendix 2", a brief discussion on each of these methods is provided.

In the Chi-square test, as there is no specified method to choose the number and size of class intervals (Johnson *et al.* 1995; Boero *et al.* 2004; Murthy *et al.* 2004), we calculate Chi-square values corresponding to six classes (<3, 3–6, 6–9, 9–12, 12–15, >15) as well as five classes (<3, 3–6, 6–10, 10–15, >15). This would, in fact, throw some light on the sensitivity of the Chi-square test. The model selection results are shown in Table 5.

Table 4

Estimated parameter values and their asymptotic standard deviations and confidence bounds

Model	Parameter values		Asymptotic standard deviation		Confidence interval (95 %)	
					Lower	Upper
Exponential	α	7.825523	σ_α	1.795298	4.306739	11.344307
Frechet	α	3.556466	σ_α	0.768384	2.050433	5.062499
	β	1.118057	σ_β	0.199992	0.726073	1.510041
Gamma	α	1.499642	σ_α	0.495271	0.528911	2.470373
	β	5.218260	σ_β	1.641766	2.000399	8.436121
Generalized exponential	α	0.175373	σ_α	0.029881	0.116806	0.233940
	β	1.700093	σ_β	0.470480	0.777952	2.622234
Inverse Gaussian	α	7.825523	σ_α	2.308987	3.299908	12.351138
	β	7.116776	σ_β	1.882572	3.426935	10.806617
Levy	α	3.727170	σ_α	1.209254	1.357032	6.097308
Lognormal	α	1.723978	σ_α	0.143150	1.443404	2.004552
	β	0.882439	σ_β	0.202445	0.485647	1.279231
Maxwell	α	5.621909	σ_α	0.526540	4.589891	6.653927
Pareto[a]	α	0.876712	σ_α	0.105778	0.669387	Capped at 0.876712
	β	0.538923	σ_β	0.150581	0.243784	0.834062
Rayleigh	α	6.885405	σ_α	0.789810	5.337377	8.433433
Weibull	α	8.559719	σ_α	1.523594	5.573475	11.545963
	β	1.356785	σ_β	0.242695	0.881103	1.832467

[a] For the Pareto distribution, we have calculated (Quandt 1966) exact standard deviations $(\sigma_\alpha, \sigma_\beta)$ of the estimated parameters

Table 5

Model selection using three criteria: the minimum Chi-square (χ^2_{value}) *criterion, the maximum log-likelihood criterion* ($\ln L$)*, and the Kolmogorov–Smirnov (K–S) minimum distance criterion*

Distribution	Minimum Chi-square		Maximum likelihood		K–S Min. distance
	χ^2_1	χ^2_2	$\ln L$	AIC	K–S distance[a]
Exponential	2.8505	1.1820	−58.0904	118.1808	0.1532 (2.3397)
Frechet	5.9953	3.5943	−59.3054	122.6108	0.1326 (11.5095)
Gamma	2.7116	1.1906	−56.9075	117.8149	0.1482 (11.5095)
Generalized exponential	3.0500	1.5111	−56.8924	117.7849	0.1532 (11.5095)
Inverse Gaussian	4.3583	2.3657	−57.4499	118.8997	0.1732 (11.5095)
Levy	10.1131	7.5517	−63.5945	129.1891	0.3224 (21.5699)
Lognormal	3.8715	1.9995	−57.3392	118.6783	0.1609 (11.5095)
Maxwell	18.6984	16.4250	−65.6992	133.3983	0.2454 (11.5095)
Pareto	16.3595	12.6241	−63.5011	131.0021	0.3055 (2.3397)
Rayleigh	10.2023	9.0612	−59.5618	121.1235	0.2507 (3.4466)
Weibull	2.7594	1.3535	−56.8651	117.7303	0.1440 (11.5095)

χ^2_1 and χ^2_2 are the Chi-square values corresponding to six classes (<3, 3–6, 6–9, 9–12, 12–15, >15) and five classes (<3, 3–6, 6–10, 10–15, >15), respectively; $\ln L$ log-likelihood value, *AIC* value of Akaike information criterion

[a] The numbers in parenthesis are the abscissa values, where the K–S distance is achieved

Table 5 shows that χ^2_1 values for the exponential (2.85), gamma (2.71), and Weibull (2.76) distributions are the least (within a tolerable limit), although the generalized exponential (3.05) is not very far away. Similarly, χ^2_2 value for the exponential (1.18) distribution is the minimum, and χ^2_2 values for the gamma (1.19), generalized exponential (1.51), and Weibull (1.35) distributions appear next to the exponential distribution. Moreover, we see that the decision from the Chi-square criterion, for the present catalog, is not very sensible for the choice of the number or size of class intervals.

On the other hand, AIC values corresponding to the gamma (117.82), generalized exponential (117.78), and Weibull (117.73) distributions appear to be the least among all AIC values. Therefore, AIC suggests the gamma, generalized exponential, and Weibull models to be the most suitable ones to represent the present earthquake catalog of northeast India and its adjoining regions. Besides, we see that the exponential (118.18) distribution has also a quite smaller AIC value.

The Frechet distribution has the minimum K–S distance (0.1326); the Weibull (0.1440), gamma (0.1482), exponential (0.1532), and generalized exponential (0.1532) distributions have also quite smaller K–S distances, giving an impression that these distributions may suitably fit the present data. To be more certain, we assess the overall matching among these distributions with the empirical distribution function by simultaneously plotting in Fig. 2.

Figure 2 reveals quite interesting facts: (i) the K–S distance value for the Frechet distribution, although it is the minimum, it fits poorly to the overall data set, (ii) the gamma, generalized exponential, and Weibull distributions match quite well to the empirical distribution, and these three distributions themselves are very close to each other, almost indistinguishable, (iii) the calculated K–S values for exponential and generalized exponentials, although they were very similar, the K–S plot suggests the generalized exponential to be more appropriate than the exponential distribution.

From the above discussion on model selection, it may be inferred that broadly three categories of distributions have emerged: the gamma, generalized exponential, and Weibull distributions have the best fitting, while exponential, Frechet, inverse Gaussian, and lognormal distributions have intermediate fitting, and the rest, namely Levy, Maxwell Pareto, and Rayleigh distributions fit poorly to the present data.

5. Conditional Probability

Having identified the suitable models in the preceding section, we now apply those models to calculate conditional probability [using (1)] of an earthquake ($M \geq 7.0$) for an elapsed time of 18 years (i.e., July 2013). These values are listed in Table 6.

Table 6 shows that the conditional probability (from the gamma, generalized exponential, and

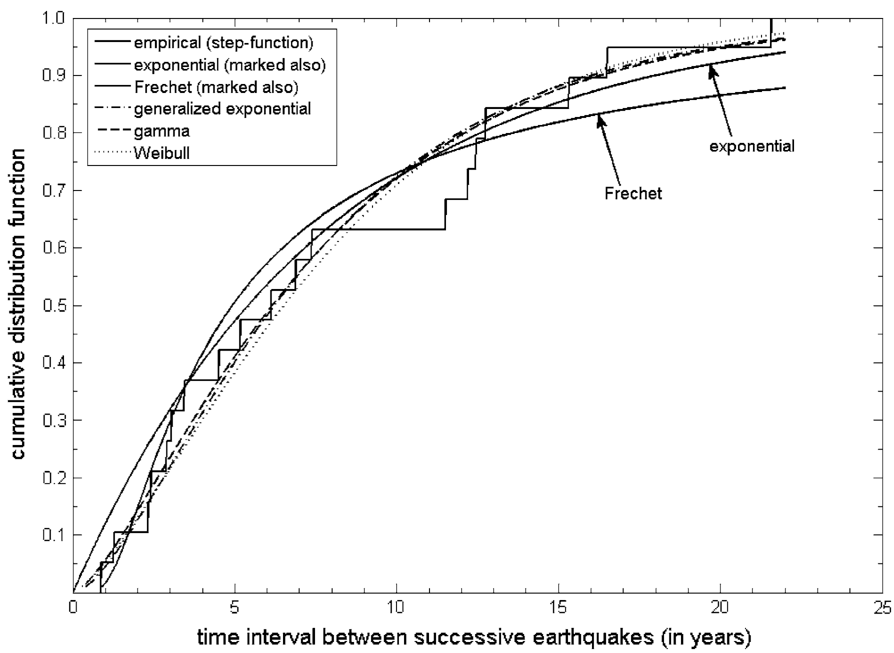

Figure 2
Empirical distribution function and the fitted distributions functions

Table 6

Estimated conditional probability for an elapsed time of 18 years (i.e., July 2013) for the study region as obtained from exponential, Frechet, gamma, generalized exponential (GE), inverse Gaussian (IG), lognormal (LN), and Weibull models

v	Year	Exponential	Frechet	Gamma	GE	IG	LN	Weibull
2	2015	0.23	0.10	0.29	0.29	0.20	0.20	0.34
5	2018	0.47	0.23	0.58	0.58	0.41	0.41	0.66
8	2021	0.64	0.32	0.75	0.75	0.57	0.56	0.83
11	2024	0.75	0.39	0.85	0.85	0.67	0.66	0.92
14	2027	0.83	0.45	0.91	0.91	0.75	0.74	0.96
17	2030	0.89	0.50	0.95	0.95	0.81	0.80	0.98
20	2033	0.92	0.55	0.97	0.97	0.86	0.84	0.99
23	2036	0.95	0.58	0.98	0.98	0.89	0.87	>0.99
26	2039	0.96	0.61	0.99	0.99	0.91	0.89	>0.99
29	2042	0.98	0.64	0.99	0.99	0.93	0.91	>0.99
32	2045	0.98	0.66	>0.99	>0.99	0.95	0.93	>0.99
35	2048	0.99	0.68	>0.99	>0.99	0.96	0.94	>0.99
38	2051	0.99	0.70	>0.99	>0.99	0.97	0.95	>0.99

Weibull distributions) of a large magnitude earthquake reaches 0.8–0.9 by 2020–2027. Besides, it may be observed that the gamma and generalized exponential distributions produce similar (up-to second decimal place) conditional probability values for the present catalog. This emphasizes the scope and suitability of the comparatively new generalized exponential distribution in seismic recurrence studies.

The conditional probability curves, generated from conditional probability values, for a combination of waiting time and elapsed time are presented in Fig. 3. These curves play a significant role in seismic zonation and microzonation, urban planning and insurance, designing of important structures such as schools, hospitals, mega-malls, and nuclear power plants, and related concerns (SSHAC 1997; Kagan and Schoenberg 2001; Baker 2008; Yadav *et al.* 2010).

6. Summary and Conclusions

Forecasting of large earthquakes, in a specified region, has been an important task for seismologists and earthquake professionals. The present research contributes to this endeavor by focusing on eleven probability distributions, namely exponential, Frechet, gamma, generalized exponential, inverse Gaussian, Levy, lognormal, Maxwell, Pareto, Rayleigh, and Weibull to analyze earthquake interevent times of large $(M \geq 7.0)$ earthquakes in the seismically active northeast India and its adjoining regions. We have briefly explained several model characteristics of these distributions, parameter estimations from the maximum likelihood method, and model selections using three goodness-of-fit methods. In addition, we have paid special attention to the problem of uncertainty measurement of the estimated model parameters. Finally, towards the end of this article, we have presented a number of conditional probability curves (also known as hazard curves) for elapsed time $\tau = 0, 5, 10, \ldots, 60$ years. These curves reveal very high seismicity in the study region.

The present study brings out the following results:

1. The gamma, generalized exponential, and Weibull distributions provide the best fitting, while exponential, Frechet, inverse Gaussian, and lognormal distributions provide intermediate fitting, and the rest, namely Levy, Maxwell Pareto, and Rayleigh distributions fit poorly to the present earthquake catalog of northeast India and its adjoining regions.
2. The conditional probability (from the gamma, generalized exponential, and Weibull distributions) of a large magnitude earthquake $(M \geq 7.0)$ in the study regions reaches 0.8–0.9 by 2020–2027.

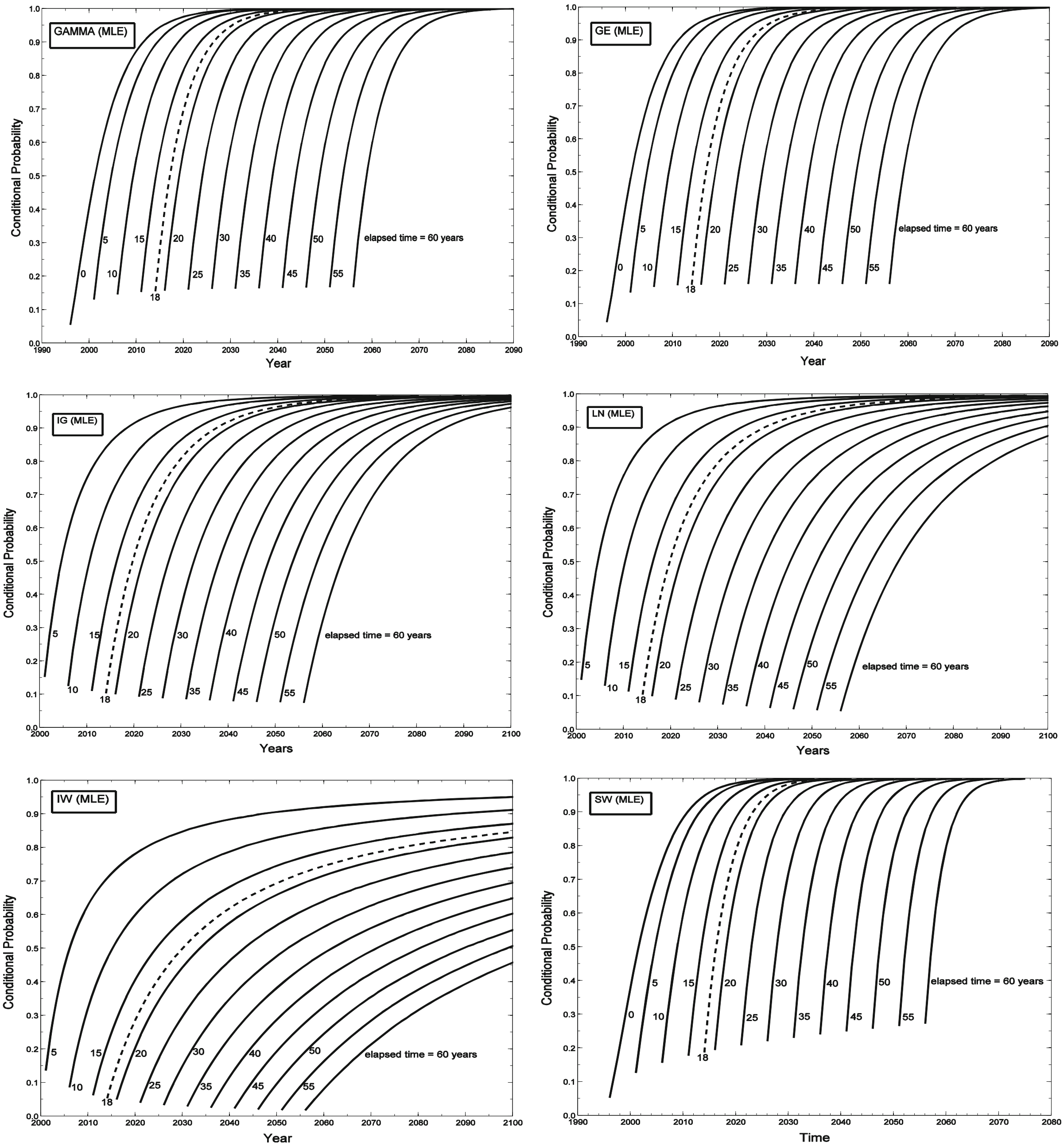

Figure 3
Conditional probability curves (hazard curves) for elapsed time $\tau = 0, 5, 10, \ldots, 60$ years, using gamma, generalized exponential (*GE*), inverse Gaussian (*IG*), lognormal (*LN*), Frechet (also called as inverse Weibull IW), and standard Weibull (*SW*) distribution for earthquake events of $M \geq 7$ in northeast India and its surrounding region. The *dot-line* represents the hazard curve corresponding to an elapsed time of 18 years, i.e., 2013

Acknowledgments

We are extremely grateful to three anonymous reviewers for their valuable comments and suggestions which helped in improving our work. We are also thankful to Dr. A. Kijko, the editor, PAAG, for his important suggestions. The first author [S.P.] acknowledges the financial support from the Council of Scientific and Industrial Research (CSIR), New Delhi.

Appendix 1

Under the standard regularity conditions (Lehmann 1998), FIM for parameter vector $\theta = (\theta_1, \theta_2)$, say, is defined as follows:

$$I(\theta) = E\left(\begin{bmatrix} \frac{\partial}{\partial\theta_1}\ln f(T;\theta) \\ \frac{\partial}{\partial\theta_2}\ln f(T;\theta) \end{bmatrix}\begin{bmatrix} \frac{\partial}{\partial\theta_1}\ln f(T;\theta) & \frac{\partial}{\partial\theta_2}\ln f(T;\theta) \end{bmatrix}\right). \quad (4)$$

The FIM can also be expressed (Nelson 1982; Hogg *et al.* 2005) as

$$I(\theta) = E\left(-\begin{bmatrix} \frac{\partial^2}{\partial\theta_1^2}\ln f(T;\theta) & \frac{\partial^2}{\partial\theta_1\partial\theta_2}\ln f(T;\theta) \\ \frac{\partial^2}{\partial\theta_2\partial\theta_1}\ln f(T;\theta) & \frac{\partial^2}{\partial\theta_2^2}\ln f(T;\theta) \end{bmatrix}_{2\times 2}\right). \quad (5)$$

Moreover, Efron and Johnstone (1990) observed that $I(\theta)$ can also be expressed in terms of the hazard function as

$$I(\theta) = E\left(\begin{bmatrix} \frac{\partial}{\partial\theta_1}\ln h(T;\theta) \\ \frac{\partial}{\partial\theta_2}\ln h(T;\theta) \end{bmatrix}\begin{bmatrix} \frac{\partial}{\partial\theta_1}\ln h(T;\theta) & \frac{\partial}{\partial\theta_2}\ln h(T;\theta) \end{bmatrix}\right). \quad (6)$$

Depending on the form of density and hazard functions of a distribution, the most convenient equation from (4, 5, 6) is used to calculate $I(\theta)$.

Appendix 2

For simplicity, we assume two competitive models F and G with density functions $f\left(x;\widehat{\theta}\right)$ and $g\left(x;\widehat{\varphi}\right)$, respectively.

The minimum Chi-square criterion is one of the conventional techniques for model selection. The Chi-square criterion consists of three steps: (i) divide the sample observations $\{t_1, t_2, \ldots, t_n\}$ into k disjoint groups of moderate length (may be equal or unequal in size) and record the observed frequencies $n_1, n_2, \ldots, n_k$(ii) compute the expected frequencies $f_1, f_2, \ldots, f_k$ and $g_1, g_2, \ldots, g_k$ based on $f\left(x;\widehat{\theta}\right)$ and $g\left(x;\widehat{\varphi}\right)$, and (iii) calculate Chi-square distances $\chi^2_{f,\text{data}}$ and $\chi^2_{g,\text{data}}$ as $\chi^2_{f,\text{data}} = \sum_{i=1}^{k}\frac{(n_i-f_i)^2}{f_i}$, $\chi^2_{g,\text{data}} = \sum_{i=1}^{k}\frac{(n_i-g_i)^2}{g_i}$. If $\chi^2_{f,\text{data}} < \chi^2_{g,\text{data}}$, then distribution F should be chosen, otherwise, G should be chosen. In the Chi-square test, the only confusion arises in selecting the number of class intervals k. There are no hard and fast rules to select the interval size (Johnson *et al.* 1995; Boero *et al.* 2004; Murthy *et al.* 2004). Thus, a reasonable number of class intervals with moderate observed frequencies are preferred in all practical applications.

The maximum likelihood criterion is entirely based on the log-likelihood values obtained in MLE. Among several competitive models, the model for which the log-likelihood value is the maximum is tagged as the best model. The maximum likelihood criterion, despite its simplicity, has a few drawbacks. For instance, it assumes that the number of parameters in each competitive model is the same. However, this presumption hardly holds true in practical situations. As a result, several modifications have been proposed over decades. Among these, AIC (Akaike 1974) has been widely used. AIC is defined as $\text{AIC} = 2k - 2\ln L$.

The Kolmogorov–Smirnov (K–S) minimum distance criterion prioritizes the competing models based on their 'closeness' to the empirical distribution function of the sample data $\{t_1, t_2, \ldots, t_n\}$. Unlike the maximum likelihood criterion, the K–S minimum distance method does not require any presumption on the number of parameters in the competitive models. Besides, the K–S test is a nonparametric and distribution free test, and; hence, it avoids the use of special tables unlike the Chi-square criterion defined previously (Johnson *et al.* 1995).

References

Akaike, H. (1974), *A new look at the statistical model identification*, IEEE Trans. Auto. Control *19*(6), 716–723.

BAK, P., CHRISTENSEN, K., DANON, L., SCANLON, T. (2002), *Unified scaling law for earthquakes*, Phys. Rev. Lett. *88* (17), 178501–178504.

BAKER, J. W. (2008), *An introduction to probabilistic seismic hazard analysis*, version 1.3; http://www.stanford.edu/~bakerjw/Publications/Baker_(2008)_Intro_to_PSHA_v1_3.pdf (retrieved on May 25, 2013).

BILHAM, R. and ENGLAND, P. (2001), *Plateau pop-up during the 1897 earthquake*, Nature *410*, 806–809.

BIS (2002), IS 1893 (part 1)-2002: *Indian standard criteria for earthquake resistant design of structures, part 1–general provisions and buildings*, Bureau of Indian Standards, New Delhi.

BOERO, G., SMITH, J. and WALLIS, K. F. (2004), *The sensitivity of Chi-squared goodness-of-fit tests to the partitioning of data*, Econometric Reviews *23*, 341–370.

DAVIS, P. M., JACKSON, D. D., and KAGAN, Y. Y. (1989), *The longer it has been since the last earthquake the longer the expected time till the next?* Bull. Seismol. Soc. Am. *79*, 1439–1456.

DIONYSIOU, D.D. and PAPADOPOULOS, G.A. (1992), *Poissonian and negative binomial modeling of earthquake time series in the Aegean area*, Phys. Earth Planet. Inter. *71*, 154–165.

EFRON, B. and JOHNSTONE, I. (1990), *Fisher information in terms of the hazard function*, Annals Statist. *18*, 38–62.

FERRAES, S.G. (2003), *The conditional probability of earthquake occurrence and the next large earthquake in Tokyo,* Japan J. Seismol. *7*, 145–153.

GSI (2000), *Seismotectonic atlas of India and its environs, Geological Survey of India*, Spec. Publ. no. 59, Kolkata.

GUPTA, H. K. and SINGH, H. N. (1986), *Seismicity of northeast India region: Part –II: earthquake swarm precursory to moderate magnitude to great earthquakes*, J. Geol. Soc. India *28*, 367–406.

GUPTA, H. K., RAJENDRAN, K., and SINGH, H. N. (1986), *Seismicity of the northeast India region: Part I: the data base*, J. Geol. Soc. India *28*, 345–365.

GUPTA, R. D. and KUNDU, D. (1999), *Generalized exponential distributions*, Aust. New Zealand J. Statist. *41*(2), 173–188.

GUPTA, R. D. and KUNDU, D. (2006), *On the comparison of Fisher information of the Weibull and GE distributions*, J. Statist. Plann. Inf. *136*, 3130–3144.

GUPTA, R. D. and KUNDU, D. (2007), *Generalized exponential distributions: existing theory and recent developments*, J. Statist. Plann. Inf. *137*(11), 3537–3547.

HOGG, R. V., MCKEAN, J. W., and CRAIG, A. T. (2005), *Introduction to mathematical statistics*, 6th Ed., PRC Press, pp. 718.

JOHNSON, N. L., KOTZ, S., and BALAKRISHNAN, N. (1995), *Continuous univariate distributions*, vol. 2, 2nd Ed., New York: Wiley.

KAGAN, Y. Y. (2007), *Inverse Gaussian distribution and its application to earthquake occurrence*; available online: http://moho.ess.ucla.edu/~kagan/igd.txt (visited on May 31 2013).

KAGAN, Y. Y. and SCHOENBERG, F. (2001), *Estimation of the upper cutoff parameter for the tapered Pareto distribution*, J. Appl. Prob. *38*, 158–175.

KAYAL, J. R. (1996), *Earthquake source process in Northeast India: A review*, Him. Geol. *17*, 53–69.

KIJKO, A. and SELLEVOLL, M. A. (1981), *Triple exponential distribution, a modified model for the occurrence of large earthquakes*, Bull. Seismol. Soc. Am. *71*, 2097–2101.

LAWLESS, J. F. (1982), *Statistical models and methods for lifetime data*, Wiley, New York.

LEHMANN, E. L. (1998), *Theory of point estimation*, Wiley, New York, pp. 616.

MATTHEWS, M. V., ELLSWORTH, W. L., and REASENBERG, P. A. (2002), *A Brownian model for recurrent earthquakes*, Bull. Seismol. Soc. Am. *92* (6), 2233–2250.

MURTHY, D. N. P., XIE M., and JIANG R. (2004), *Weibull models*, John Wiley and Sons, New Jersey, 1st Ed., pp. 383.

NANDY, D. R. (1986), *Tectonics, seismicity and gravity of North eastern India and adjoining region*, Geol. Surv. India Memoir. *119*, 13–16.

NANDY, D. R. (2001), *Geodynamics of north eastern India and the adjoining region*, 1st Ed, ACB Publications, Kolkata, pp. 209.

NELSON, W. (1982), *Applied life data analysis*, Wiley, New York.

PAPAZACHOS, B. C., PAPADIMITRIOU, E. E., KIRATZI, A. A., PAPAIOANNOU, C. A., KARAKAISIS, G. F. (1987), *Probabilities of occurrence of large earthquakes in the Aegean and surrounding area during the period 1986–2006*, Phys. Earth Planet Inter. *125*(4), 597–612.

PARVEZ, I. A. and RAM, A. (1997), *Probabilistic Assessment of earthquake hazards in the north-east Indian Peninsula and Hindukush regions*, Pure Appl. Geophys. *149*, 731–746.

PASARI, S. and DIKSHIT, O. (2013), *Impact of three-parameter Weibull models in probabilistic assessment of earthquake hazards*, Pure Appl. Geophys. doi:10.1007/s00024-013-0704-8.

PISARENKO, V. E., SORNETTE, D., RODKIN, M. V. (2010), *Distribution of maximum earthquake magnitudes in future time intervals: application to the seismicity of Japan (1923–2007)*, Earth Planets Space *62*, 567–578.

QUANDT, R. E. (1966), *Old and new methods of estimation and the Pareto distribution*, Metrika *10*, 55–82.

REID, H. F. (1910), *The mechanics of the earthquake, the California earthquake of April 18, 1906*, Report of the State Investigation Commission, Carnegie Institution of Washington, Washington, vol. 2.

SORNETTE, D. and KNOPOFF, L. (1997), *The paradox of the expected time until the next earthquake*, Bull. Seism. Soc. Am. *87*, 789–79.

SOTOLONGO-COSTA, O., ANTORANZ, J. C., POSADAS, A., VIDAL, F., and VAZQUEZ, A. (2000), *Levy flights and earthquakes*, Geophys. Res. Let. *27*(13), 1965–1968.

SSHAC (Senior Seismic Hazard Analysis Committee), *Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts* (1997), US Nuclear Regulatory Commission Report, CR-6372, Washington, DC, p. 888.

THINGBAIJAM, K. K. S., NATH, S. K., YADAV, A., RAJ, A., WALLING, M. Y., and MOHANTY, W. K. (2008), Recent seismicity in northeast India and its adjoining region, J. Seismol. *12*, 107–123.

YADAV, R. B. S., TRIPATHI, J. N., RASTOGI, B. K., DAS, M. C., and CHOPRA, S. (2010), *Probabilistic assessment of earthquake recurrence in northeast India and adjoining regions*, Pure Appl. Geophys. *167*, 1331–1342.

YAZDANI, A. and KOWSARI, M. (2011), *Statistical prediction of the sequence of large earthquakes in Iran*, IJE Trans. B: Appl. *24*(4), 325–336.

(Received July 28, 2013, revised December 21, 2013, accepted January 8, 2014, Published online February 27, 2014)