Nat Hazards (2011) 59:553–570
DOI 10.1007/s11069-011-9776-x

ORIGINAL PAPER

The distribution of annual maximum earthquake magnitude around Taiwan and its application in the estimation of catastrophic earthquake recurrence probability

Jui-Pin Wang · Chung-Han Chan · Yih-Min Wu

Received: 11 August 2010 / Accepted: 4 March 2011 / Published online: 17 March 2011

Abstract The annual maximum earthquake magnitudes around Taiwan from 1900 to 2009 are presented in this paper. Using the distribution of the AMEM, a probabilistic framework to estimate the recurrence probability of a large-size earthquake is also proposed and an illustration was made in this paper. The mean value of the 110-AMEM is 6.433, and the coefficient of variation is around 10%. The results of two goodness-of-fit tests show that the Gamma and lognormal distributions are relatively suitable to represent the AMEM around Taiwan among five common probability distributions. Using the proposed approach, the recurrence probability is 4% for an earthquake with magnitude greater than 7.5 in a 1-year period around Taiwan. More site-specifically, the probability is around 5% in Central Taiwan for such an earthquake to occur in a 50-year period.

Keywords Annual maximum earthquake magnitude · Goodness-of-fit test · Lognormal distribution · Gamma distribution

1 Introduction

Taiwan is located at the plate boundary where the arc-continent collision occurs between the Luzon volcanic arc at the west edge of the Philippine Sea plate and the Chinese continental margin of the Eurasian Plate (Suppe 1984). The tectonic setting has produced the island of Taiwan where 23 million people reside as of today. As a result of the regional tectonic movements, most of Taiwan is under a northwest–southeast compression at a convergence rate of about 8 cm/year (Yu et al. 1997). The Taiwan orogeny, starting around

J.-P. Wang (✉)
Department of Civil & Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
e-mail: jpwang@ust.hk

C.-H. Chan · Y.-M. Wu
Department of Geosciences, National Taiwan University, Taipei, Taiwan

4 Ma (Suppe 1984), is relatively young in the geologic timescale. The island has a high rate of crustal deformation and strong seismic activity. The seismic activities have produced a few disasters onto the Taiwanese people (Wang and Shin 1998; Wu et al. 2008).

The scientific community has spent a large effort in trying to forecast the occurrence of earthquakes, especially the catastrophic (large size) ones (Jordan 2006; Field 2007). However, current scientific knowledge seems insufficient to resolve three issues, i.e., where, when, and what is the magnitude of a large earthquake. This leads to the inevitable uncertainty in the earthquake forecast. In addition, the nature of randomness increases the difficulty in developing a reliable forecast. This combination of epistemic (imperfect knowledge) and aleatory (random nature) uncertainties explains disagreements on the most suitable model to forecast earthquake occurrences. The probabilistic seismic hazard analysis, PSHA, (Cornell 1968) is one of the methodologies that has gained its popularity and has been adopted to estimate the site-specific seismic hazard in terms of annual rates of ground accelerations (US Nuclear Regulatory Commission 2007). In general, the framework of the PSHA comprises the probabilistic analysis based on distribution and variation of the seismicity record around the site of interest. However, the legitimacy of PSHAs has been debated in the scientific community (Castanos and Lomnitz 2002; Krintzsky 2002; Bommer and Crowley 2006).

An estimate of maximum earthquake magnitude, $m_{\max}$, is needed in many engineering applications, but it is surprising how little research has targeted on developing the appropriate methodology for its estimation (Kijko 2004). PSHA, for instance, requires determining the maximum earthquake magnitude of a seismic zone in order to conduct seismic hazard computation. To the authors' best knowledge, the logic tree technique is used in most PSHA practices, by which the uncertainty of maximum magnitudes can be addressed/controlled using multiple maximum magnitudes with respective weighting. It must be noted that the maximum magnitude and corresponding weighting are interpreted in a deterministic manner based on available investigation and research.

One of the approaches to estimate maximum magnitude was based on the empirical relationship between the maximum earthquake magnitude and tectonic quantities (e.g., fault dimension) investigated for the specific region of interest (Papastamatiou 1980; Jin and Aki 1988; Wells and Coppersmith 1994; Anderson et al. 1996). However, a shortcoming inherited in the empirical estimation is the significant uncertainty (Kijko 2004). Probabilistic approaches conducting the statistical analysis on the historical seismic data, say earthquake catalog, were also developed in estimating $m_{\max}$. Pisarenko et al. (1996) employed a probabilistic framework to estimate the possible maximum earthquake magnitude in California and Italy. Kijko (2004) proposed a generic equation considering the adopted statistical model and available observational records.

In this study, a statistical procedure was employed to determine suitable probability models to estimate the distribution of the annual maximum earthquake magnitude (AMEM; denoted as $\tilde{M}$) and its model parameters. The AMEM is defined as the maximum magnitude in a single year extracted from the earthquake catalog for the specific region. A method that utilizes the AMEM distribution to estimate the recurrence probability of large earthquakes is proposed in this paper. An illustration was made by using the earthquake catalog around Taiwan since 1900 by the Central Weather Bureau (Wang and Shin 1998; Wu et al. 2008), and the recurrence probability of large-size earthquakes in the region was also evaluated.

2 Methodology

Firstly, five common probability distributions were employed to model the AMEM, and their suitability was examined by two different statistical goodness-of-fit tests, Kolmogorov–Smirnov (K–S) and Chi-square tests. Once the probability models were found acceptable, the distribution of the AMEM can be forecasted. As a result, the recurrence probability of large-size earthquakes in a yearly basis can be estimated by using the probability distribution of the AMEM. The proposed methodology was illustrated using the earthquake catalog for Taiwan.

2.1 Earthquake catalog

The earthquake catalog from 1900 to date around Taiwan was used in this study (Wang and Shin 1998; Wu et al. 2008). Until the end of 2009, the catalog approximately recorded 66,000 events with $M \geq 3.0$. Note that the magnitude in the catalog is in the local magnitude (M_L) scale. Figure 1 displays the spatial distribution of the events. Figure 2 shows the number of events and the minimum magnitude that was recorded in each year. It was found that the catalog can be characterized into three periods. From 1900 to 1933, the minimum magnitude was at least 4.0 but the variability was relatively significant. In this period, the number of events was very small. From 1934 to 1972, the minimum magnitude was constantly kept around 4.0 and the number of events increased compared to the

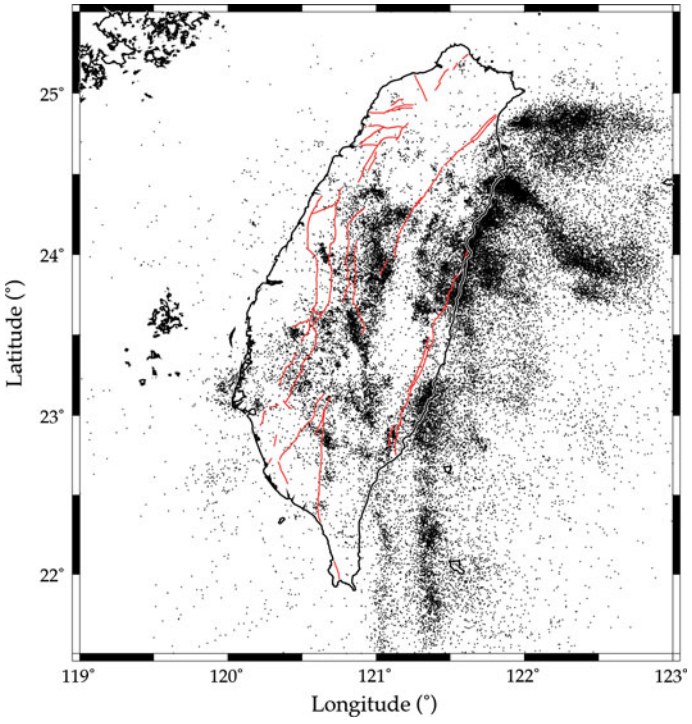

Fig. 1 Spatial distribution of the earthquakes recorded since 1900 around Taiwan; note that *solid lines* and *dot points* represent *the fault lines* and earthquake occurrences, respectively

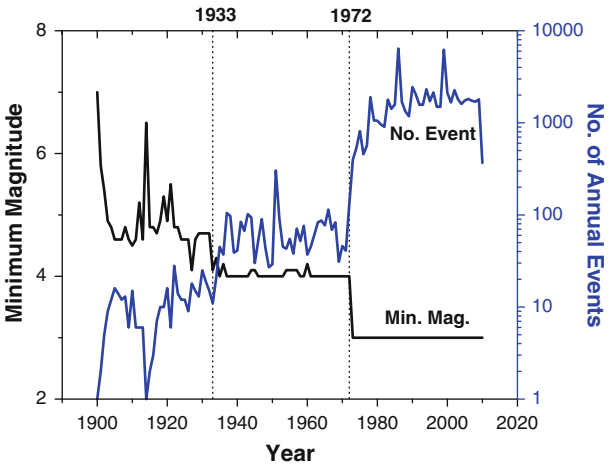

Fig. 2 Distribution of the number of annual events and minimum magnitude (recorded threshold magnitude) from 1900 to 2009 around Taiwan

previous period. The last period started at 1973. The minimum magnitude was equal to 3.0 and the number of events increased dramatically. The causes to those observations were: (1) the high minimum magnitude that filtered out most small-scale earthquakes and (2) the limited capability of early instrumentation in the detection of earthquake that accounted to the incompleteness.

As indicated by Fig. 2, incompleteness is inevitable in the earthquake catalog, especially in the early period or before the employment of modern earthquake detection techniques. In this study, the AMEM is the parameter of interest and the effect of the incompleteness could be insignificant because large-size earthquakes were easily detected or recorded in human history. However, it was expected that the accuracy of the AMEM in the early period might not as satisfying as that with the use of modern instrumentation. The uncertain AMEM in the early era could result from the limitations of early instrumentation and interpretations on descriptive information. To examine their uncertainties or reliability, further studies are suggested, but this is not the major objective of this study.

In addition, the 110-AMEM extracted from the original catalog can be considered a time-serious datum. Its dependency and correlation were examined by performing the sample autocorrelation function (ACF, $\hat{\rho}(h)$) which can be expressed as:

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \tag{1.1}$$

where h is lag (from 0 to 109 years in this case); $\hat{\gamma}(h)$ is the sample autocovariance function defined as:

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n \tag{1.2}$$

where n is the number of samples (equal to 110 in this case) and $\bar{x}$is the mean of the 110-AMEM. When the sample ACF is below the corresponding boundaries associated with the sample characteristics and no substantial decay and periodic component are found, the sample can be treated as independently and identically distributed in time, or the time series is considered stationary.

2.2 Probable probability models and goodness-of-fit tests

After the series of AMEM was obtained, probability models were used to simulate the AMEM distribution. Without prior information regarding which models can satisfyingly simulate the AMEM distribution, five models were selected in this study, which were (1) normal distribution, (2) lognormal distribution, (3) Gamma distribution, (4) Beta distribution, and (5) Pareto distribution. Next, two goodness-of-fit tests were used to examine their respective suitableness. It must be noted that the model parameters were converted from the statistics (e.g., mean, standard deviation) of the 110-AMEM without performing further calibration. In addition, the modal parameters A and B in the Beta distribution were selected as the minimum and maximum among the 110-AMEM, respectively, instead of performing further calibration to obtain the parameters.

The Kolmogorov–Smirnov (K–S) and Chi-Square tests are two of the widely used goodness-of-fit tests in statistics (Ang and Tang 2003). The framework of the K–S test is that the maximum difference between the observed and theoretical cumulative probabilities needs to be less than a critical value associated with the size of the observational samples and the level of significance (α). For a sample size n, the observed cumulative probability $(S_n(x))$ can be determined as:

$$S_n(x) = \begin{cases} 0 & x < x_1 \\ k/n & x_k \le x < x_{k+1} \\ 1 & x \ge x_n \end{cases} \tag{2}$$

where x_k = the kth observed value sorted in an ascending order. Let $F_X(x)$ be the theoretical cumulative distribution function for a given probability model; then, the maximum difference, D_n, between $S_n(x)$ and $F_X(x)$ over the entire range can be expressed as:

$$D_n = \max|F_X(x) - |S_n(x) \tag{3}$$

D_n can be treated as a random variable, and for a given level of significance, the probabilistic relationship between D_n and critical value $\left(D_n^{\alpha}\right)$ can be expressed as:

$$\mathbf{P}\left(D_n \le D_n^{\alpha}\right) = 1 - \alpha \tag{4}$$

The assumed model is acceptable if D_n is less than the critical value. In this study, a level of significance of 0.05 was employed.

On the other hand, the Chi-square goodness-of-fit compares the observed frequencies (n_i) and theoretical frequencies (e_i) computed by a given probability model. As the data are sorted into k-intervals, the quantity, $\sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i}$, will approach the Chi-square distribution. As a result, if the quantity is less than the critical value following the Chi-square distribution at the cumulative probability of (1–α), the adopted probability model is acceptable at the level of significance. As the K–S test, the level of significance of 0.05 was employed in this study.

2.3 Evaluation of recurrence probability of large-size earthquakes

After the suitability of the five models was examined, they were used to estimate the recurrence probabilities of large-size earthquakes. If the AMEM $\left(\tilde{M}\right)$ can be portrayed acceptably by a specific probability model and its cumulative distribution function (CDF) is denoted as $F_{\tilde{M}}(\tilde{m})$, the probability (denoted as **P**) of $\tilde{M} > m*$ can be estimated as:

$$\mathbf{P}(\tilde{M} > m*) = 1 - \mathbf{P}(\tilde{M} \leq m*) = 1 - F_{\tilde{M}}(m*) \tag{5}$$

It was expected that for a man-made structure with a long design life, it is more likely to encounter a large-size earthquake during its service life. Therefore, it was appropriate to estimate the maximum magnitude for engineering design considering the life of the structure. If the occurrence of the AMEM in every year was independent, the probability that at least one AMEM is greater than $m*$ within a $t*$-year period can be expressed as:

$$\mathbf{P}(\tilde{M} > m*, \textbf{ at least once in } t * \textbf{ years}) = 1 - [F_{\tilde{M}}(m*)]^{t*} \tag{6}$$

Note that as a 1-year period was considered ($t* = 1$), Eq. 5 was identical to Eq. 6, both representing the probability of the AMEM greater than a given $m*$ in a 1-year period.

It must be noted that Eq. 6 represents the probability of at least one AMEM greater than $m*$ for the $t*$-year period considered, denoted as $\mathbf{P}\{\tilde{M} > m* \text{ in } t* \text{ year}\}$. Let $\mathbf{P}\{M > m* \text{ in } t* \text{ year}\}$ denote the probability of at least one earthquake occurrence with magnitude $M > m*$ in $t*$-years, which can be regarded as the recurrence probability. Considering that a large-size earthquake was the AMEM at its occurrence, the $\mathbf{P}\{\tilde{M} > m* \text{ in } t* \text{ year}\}$ can be used to estimate $\mathbf{P}\{M > m* \text{ in } t* \text{ year}\}$. Take $m* = 7.9$, for example, the number of AMEM > 7.9 out of 110 AMEM observations was one, equal to one earthquake magnitude >7.9 out of 66,000 earthquake observations. As a result, the same number of such an event that was observed accounted to the same recurrence rate. On the contrary, the number of AMEM > 4.0 out of 110 observations was 110, which can be expected to be much less than the number of earthquake with magnitude >4.0 out of 66,000 observations. As a result, the proposed approach is more applicable for estimating the recurrence probability for catastrophic earthquakes ($\mathbf{P}\{M > m* \text{ in } t* \text{ year}\}$), say $M > 7.0$.

2.4 Analytical tool

The techniques of using the Excel Spreadsheet incorporated with Visual Basic Application (VBA) are used to carry out the aforementioned analysis. An excel file was developed in-house specifically to preprocess the catalog, to extract AMEM, and to perform statistical analyses (K–S tests, Chi-square test, sample ACF) and numerical computations (estimation of recurrence probabilities).

3 Results

3.1 Characteristics of AMEM around Taiwan

Figure 3 shows the scatter plot and histogram of the 110 observations of AMEM from 1900 to 2009. Table 1 summarizes the statistics including mean, standard deviation, Pearson coefficient, mean of the logarithm of AMEM, etc. The minimum and maximum AMEM were 5 and 8, respectively; the mean value and standard deviation of AMEM were 6.433 and 0.599, respectively, producing a coefficient of variation of 10% approximately. The Pearson coefficient of 0.665 indicated that the distribution of AMEM featured skewness to the right. The modal parameters converted from the statistics are also shown in Table 1. The α and β values for the Gamma distribution were 115.45 and 0.056, respectively; the x_m and α values for the Pareto distribution were 11.80 and 5.89, respectively. For the Beta distribution, the model modal parameters, α and β, were 2.51 and 2.75, respectively.

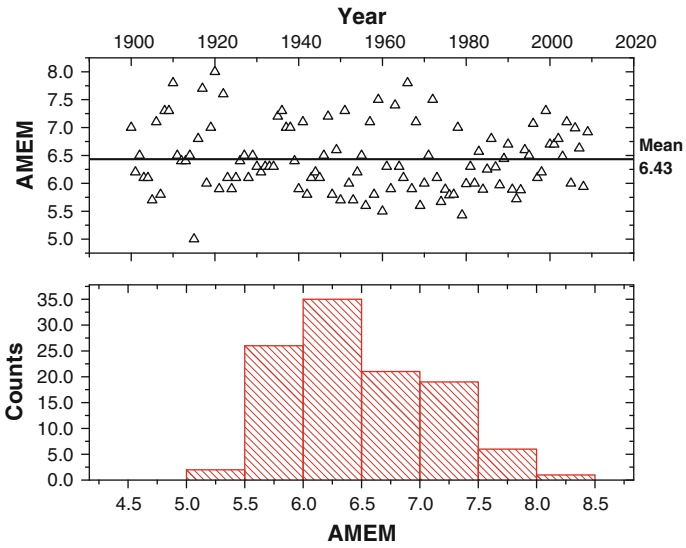

Fig. 3 Scatter plot and histogram of the AMEM from 1900 to 2009 around Taiwan

Table 1 The statistics of the annual earthquake maximum magnitude from 1900 to 2009 around Taiwan

Parameters	Values
Mean	6.433
Standard deviation (SD)	0.599
Coefficient of variation [%]	9.307
Minimum	5.000
Maximum	8.000
Median	6.300
Pearson coefficient	0.665
Mean of Logarithm of magnitude (λ)	0.807
SD of Logarithm of magnitude (ξ)	0.040
α, β for Gamma distribution	115.45, 0.056
x_m and α for Pareto distribution	11.8, 5.89
α, β for Beta distribution	2.51, 2.75

Figure 4 shows the mean and coefficient of variation (C.O.V.) of the AMEM for different periods. It indicates that the mean value was pretty uniform, but the level of variation slightly depended on the period that was considered. Higher dispersion was observed in 1900–1919 and 1960–1979. The level of variation decreased after 1980 till now. One of the probable explanations is that the observations obtained by the modern instrumentation are more accurate and reliable; in other words, the aforementioned factors (e.g., early instrumentation limitations, interpretations on descriptive data) might have an impact on the data reliability leading to the high variation found in the first period (1900–1919) in Fig. 4. Another cause is the natural randomness that resulted in a high dispersion in 1960–1979.

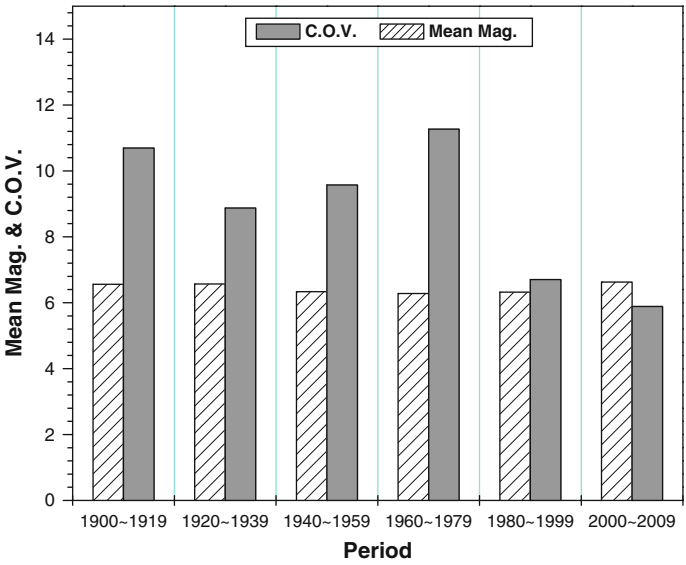

Fig. 4 Mean and C.O.V. of the AMEM around Taiwan at different periods since 1900

Figure 5a shows the sample ACF that is close to zero for all nonzero lags (1–109) and it is within 95% boundaries ($\pm 1.96/\sqrt{n}$) (n is sample size equal to 110). Figure 5b shows the absolute ACF and it indicates that no substantial decay and periodic component were found. It verifies that the AMEM is independently and identically distributed and does not demonstrate a substantial periodicity.

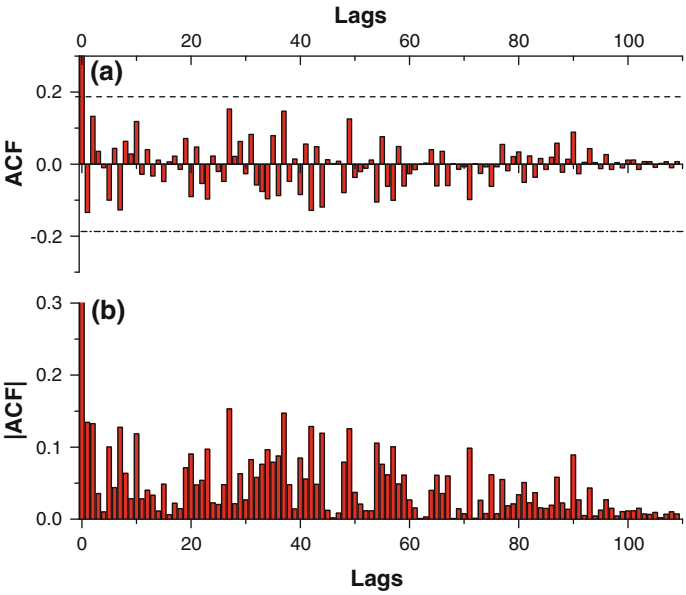

Fig. 5 Sample autocorrelation function (ACF) of the AMEM since 1900 around Taiwan

3.2 Results of goodness-of-fit tests

The comparison of the theoretical and the observed (using the K–S test) cumulative probabilities for the five distributions is shown in Fig. 6. Table 2 summarizes the maximum difference (D_n) and corresponding magnitudes using the five distributions. It was found that the degree of goodness-of-fit to simulate the AMEM was the lognormal distribution followed by Gamma, normal, Beta, and Pareto distributions. However, it must be noted that D_n was very close between lognormal (0.0938) and Gamma (0.0977) distributions. Except the Pareto distribution, the four models were suitable based on the K–S test at the significance level of 5% corresponding to a critical value $\left(D_n^{\alpha}\right)$ of 0.13.

The observed frequencies and theoretical frequencies computed by the five probability distributions are shown in Fig. 7. Note that the bin width of 0.1 was selected and 31 intervals were presented accordingly. The Chi-square quantities for normal, lognormal, and Gamma distributions were 41.1, 38.6, and 38.5, respectively, which were less than the critical value of 41.3 at the level of significance of 0.05. As a result, the three models were acceptable based on the Chi-square test. Since the theoretical frequencies at some magnitude bins were zero computed by the Beta and Pareto distributions, the Chi-square quantities cannot be computed due to the denominator as zero. In an attempt to compute

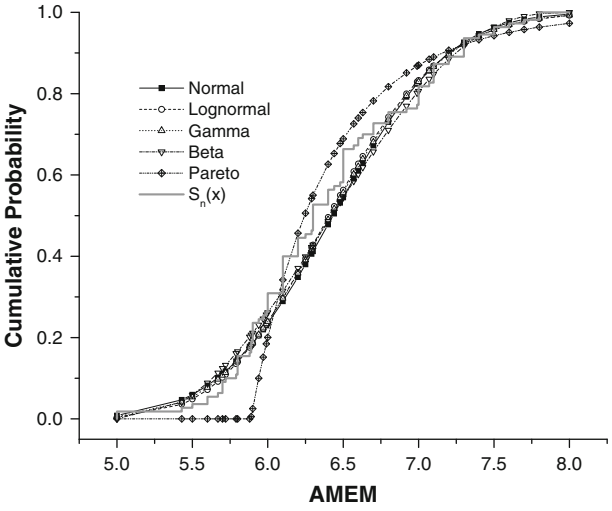

Fig. 6 Comparison of the observed and five theoretical cumulative probabilities using the K–S test

Table 2 Summary of two goodness-of-fit tests

Models	Normal	Lognormal	Gamma	Beta	Pareto
D_n	0.110	0.094	0.098	0.110	0.202
@ Magnitude	6.5	6.1	6.5	6.5	5.9
Suitability (Rank) in terms of D_n	Yes (3)	Yes (1)	Yes (2)	Yes (4)	No (5)
χ^2	41.1	38.6	38.5	51.2	63.1
Suitability (Rank) in terms of χ^2	Yes (3)	Yes (2)	Yes (1)	No (4)	No (5)

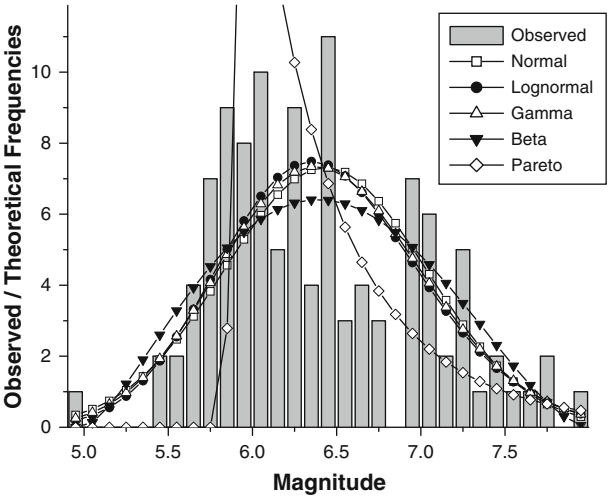

Fig. 7 Comparison of the observed and five theoretical frequencies in terms of magnitude for the 110-AMEM

the Chi-square quantities for the Beta and Pareto distributions, the Chi-square quantities were computed in the manner that only the intervals with nonzero theoretical frequencies were summed, which would generate underestimated Chi-square quantities. Using the approach, the Chi-square quantities for the Beta and Pareto distributions were 51.2–63.1, respectively, still greater than the critical value, which indicated that both distributions were not proper to simulate the AMEM. The summary of Chi-square test results is also shown in Table 2.

According to the result of the goodness-of-fit, both tests indicated that the normal, lognormal, and Gamma distributions are acceptable among the five probability models. One different conclusion between the two tests is that the Beta distribution is considered suitable using the K–S test but it is not appropriate in the Chi-square test. The other minor difference is that the lognormal distribution is relatively suitable than the Gamma distribution according to the K–S test, but it is opposite based on the Chi-square test. With very close statistical quantities in both tests, it indicates that the two distributions are equally capable of portraying the AMEM. In addition, both tests show that the Pareto distribution was not acceptable to simulate the AMEM.

3.3 Recurrence probabilities of a large-size earthquake

As mentioned in Sect. 3.2, the normal, lognormal, and Gamma distributions are acceptable to simulate the AMEM based on two goodness-of-fit tests. They were adopted to estimate the recurrence probability of a large-size earthquake using Eqs. 5 and 6. Figure 8 shows the recurrence probability considering four different periods that are 1, 10, 50, and 100 years. It was found that the patterns were similar with the fact that the recurrence probability in the high magnitude range (say $M > 7.5$) was lower using the normal distribution. According to the result, the recurrence probabilities of an earthquake with magnitude greater than 7.5 are 0.037, 0.043, and 0.042 by using the normal, lognormal, and Gamma distributions, respectively. It indicates approximately a 4% probability for an

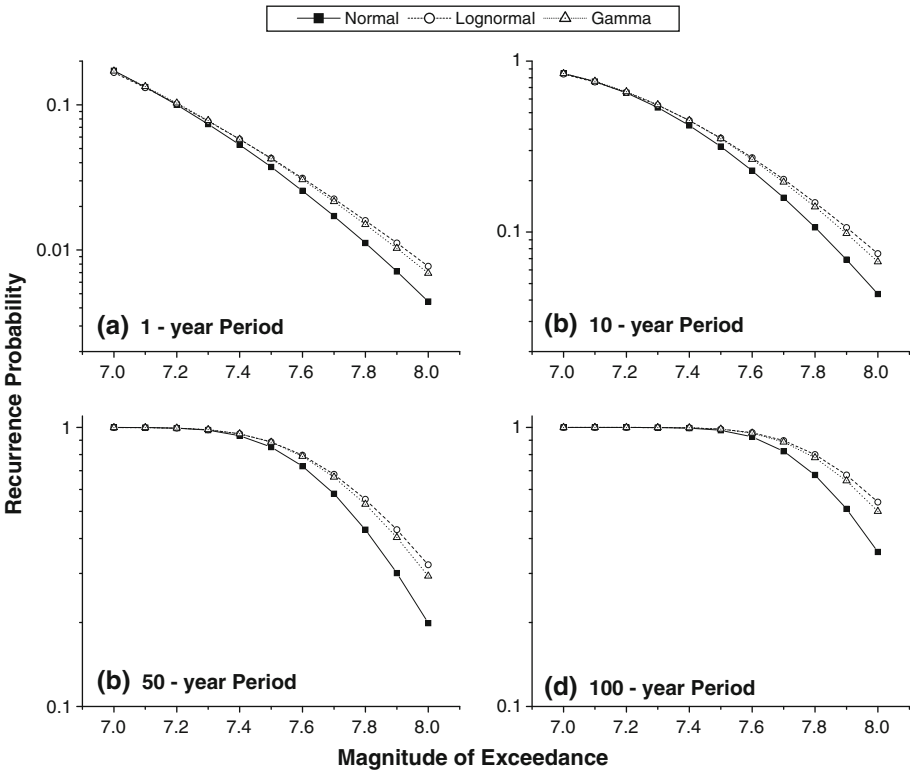

Fig. 8 Recurrence probabilities estimated by the proposed approach using the three acceptable models in representing the AMEM around Taiwan, **a** 1-year period, **b** 10-year period, **c** 50-year period, and **d** 100-year period

earthquake with magnitude >7.5 around Taiwan every year, corresponding to a return period of 25 years.

4 Discussion

4.1 Model uncertainty and best-estimate recurrence probabilities

In this study, the five widely used probability distributions are used in an attempt to portray the AMEM. Three of them are considered statistically appropriate, and the recurrence probabilities are then estimated accordingly. This finding raises the legitimate questions: are there more suitable probability distributions capable of portraying the AMEM and what is the basis to preselect the five models? We agree that it is probable to find other models or to calibrate modal parameters of the five models to improve the simulation on the specific AMEM data in this study. However, new models, if there is any, also cast the uncertainty since it is believed that not a probability distribution is capable of perfectly portraying the randomness of a variable. The better fitting to the data is simply a numerical coincidence, and it also lacks the physical support in the correlation between the model and variable. In this study, the five models feature a wide application in engineering and sciences and their

parameters can be obtained easily from the statistics (i.e., mean, standard deviation) without tedious calibration (e.g., Weibull distribution). This becomes the motivation of selecting the five models in this study.

Based on the goodness-of-fit test, the relative suitableness of the probability models was determined, which reveals that lognormal and Gamma distributions are relatively appropriate followed by the normal distribution. The Beta and Pareto distributions are regarded to be unacceptable based on the statistical tests. In order to obtain the best-estimate recurrence probability ($\mathbf{P}_{\mathrm{BE}}$), the logic tree technique was adopted by weighting the recurrence probabilities estimated by a probability model. Therefore, $\mathbf{P}_{\mathrm{BE}}$ can be expressed as:

$$\mathbf{P}_{\mathrm{BE}} = \sum_{i=1}^{k} w_i \mathbf{P}_i \tag{7}$$

where w_i and $\mathbf{P}_i$ denote the weight and recurrence probability computed by Eq. 6 for a given probability model, respectively. According to the statistical tests, the weights of 0.4, 0.4, and 0.2 were assigned for the lognormal, Gamma, and normal distributions in this study. Figure 9 shows the best-estimate recurrence probability of large-size earthquakes. The recurrence probabilities of an earthquake with $M > 7.5$ are 4, 35, 88, and 99% considering a 1-, 10-, 50-, and 100-year period, respectively. The result reinforces and corresponds to the agreement that Taiwan is located in a seismically active region.

4.2 Uncertainty of recurrence location

A prediction model that ignores the uncertainty of recurrence locations would limit its applicability on earthquake engineering, seismic hazard mitigation, etc. Unfortunately, the proposed methodology falls within this category although its analytical framework was found to be legitimate. Two logical methods are introduced herein to simulate the uncertainty of recurrence locations, and they are incorporated into the proposed methodology to develop a site-specific catastrophic recurrence probability forecast model.

In the first approach, a "probable zone" is introduced to address the uncertainty of recurrence locations. Then, similar to a PSHA, a uniform distribution is adopted,

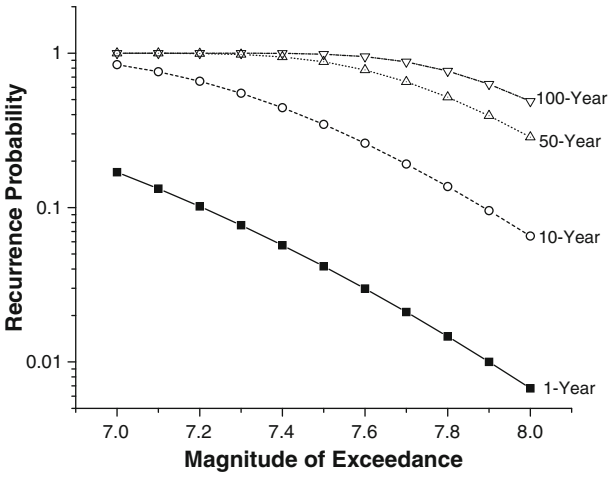

Fig. 9 Best-estimate recurrence probability considering the model uncertainty

considering that the probability of the AMEM occurring at any location within the zone is equal. Figures 10 and 11 illustrate two schemes in the lineation of the probable zones around Taiwan based on the 110 locations where the AMEM was observed. In Fig. 10, one, broad probable zone encircling all observational locations was sketched. The region outside the probable zone was considered as the background zone. In contrast, Fig. 11 displays four probable zones, and the region not belonging to any of them was also regarded as the background zone. As a result, the probability of an AMEM occurring inside a specific probable zone ($\mathbf{P}_{Zi}$) was:

$$\mathbf{P}_{Zi} = \frac{n_{Zi}}{N} \tag{8}$$

where n_{Zi} = number of observations inside the probable zone Z_i; N = the total number of observations. Combining Eqs. 7 and 8, the best-estimate, site-specific recurrence probability ($\bar{\mathbf{P}}$) can be expressed as:

$$\bar{\mathbf{P}} = \mathbf{P}_{Zi} \times \mathbf{P}_{BE} \tag{9}$$

Note that $\bar{\mathbf{P}}$ for the background zone estimated by Eq. 9 might approach zero because there could be no observations in the zone such as the case shown in Fig. 10. To account for the

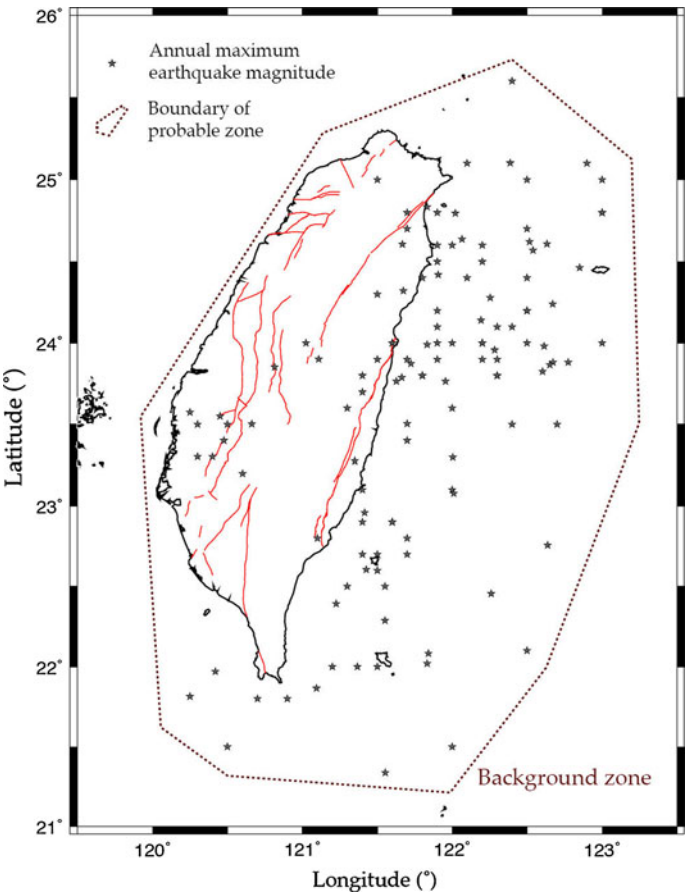

Fig. 10 One scheme to consider the uncertainty of recurrence locations by proposing a broad probable zone

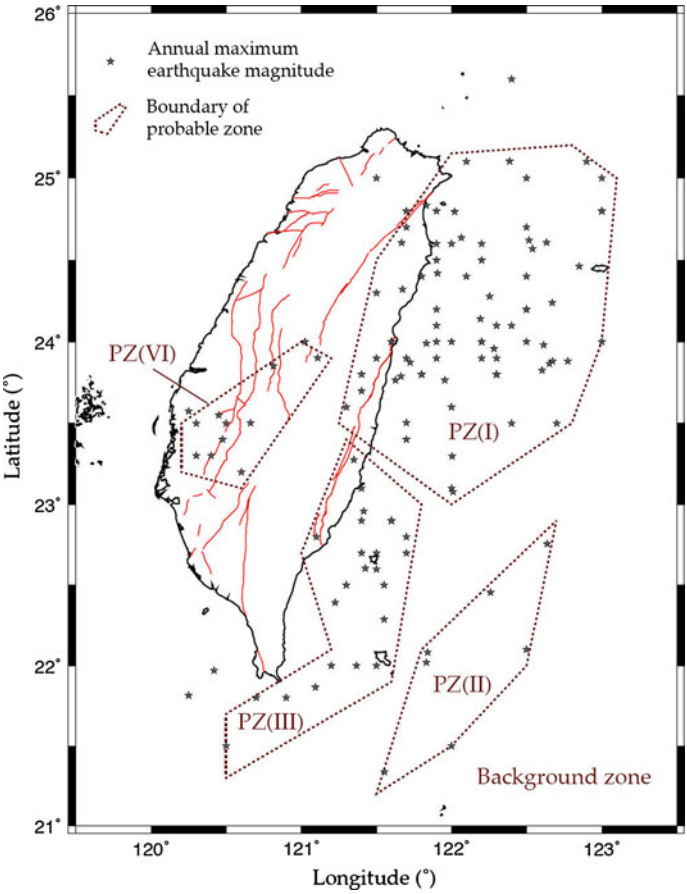

Fig. 11 The other scheme to consider the uncertainty of recurrence locations by proposing a few probable zones

uncertainty in location and to provide reasonable conservatism, the probability of the occurrence at the background ($\mathbf{P}_{\mathrm{BKZ}}$) was introduced and the best-estimate, site-specific recurrence probability became:

$$\bar{\mathbf{P}} = \max[\mathbf{P}_{\mathrm{Z}i}, \mathbf{P}_{\mathrm{BKZ}}] \times \mathbf{P}_{BE} \tag{10}$$

Using the proposed approach, the site-specific, best-estimate recurrence probabilities were evaluated. In Fig. 11, $\mathbf{P}_{\mathrm{Z1}}$, $\mathbf{P}_{\mathrm{Z2}}$,…, $\mathbf{P}_{\mathrm{Z5}}$ were 0.632, 0.056, 0.184, 0.096, and 0.032, respectively. In such a case, the distribution of the recurrence probability for a 50-year period for the five zones is shown in Fig. 12. It was found that the pattern is similar as expected due to Eq. 10 in which the second term in the right side is independent of the probable zone. It also indicates that for Central Taiwan, i.e., PZ(II), the recurrence probability of an earthquake with magnitude >7.5 is 5% for a 50-year period, which is more realistic based on the available observations.

Using this approach, the determination of the probable zones was subjective and as a result, the uncertainty was inevitable and the estimates of the site-specific recurrence

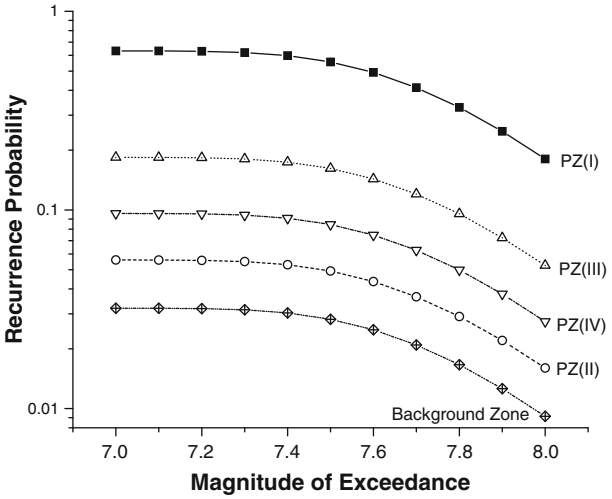

Fig. 12 Site-specific recurrence probability distribution considering a 50-year period within for the five probable zones in Fig. 11

probability could differ when different individuals perform the analysis. To improve the reliability of the approach, more geological evidence, e.g., active fault locations, plate boundaries, etc., can be used as critical references to decide the best-estimate probable zones. It is expected that the uncertainty regarding the recurrence location cannot be addressed perfectly in this framework so that an alternative approach is proposed and it is elaborated in the next paragraph.

In the second approach, the number (n_r) of observations within a given distance (forming a circular probable zone) from the given location of interest can be counted as illustrated in Fig. 13. Similar to the first approach, the best-estimate, site-specific recurrence probability for the location can be estimated as:

$$\bar{\mathbf{P}} = \max\left[\frac{n_r}{N}, \mathbf{P}_{\mathrm{BKZ}}\right] \times \mathbf{P}_{\mathrm{BE}} \tag{11}$$

Taking two locations in Fig. 13 for example, when the radius is 100 km, n_r/N for the location (122°E, 24°N) and (121°E, 22°N) is 0.552 and 0.168, respectively. Therefore, $\bar{\mathbf{P}}$ of the two locations estimated using Eq. 11 is shown in Fig. 14a. In this approach, the selected radius would have a significant impact on recurrence probability. As shown in Fig. 14b, when the radius was lowered to 50 km, for the location (122°E, 24°N) and (121°E, 22°N), the recurrence probabilities decreased to 0.248 and 0.048, respectively. As a result, $\bar{\mathbf{P}}$ followed a 55% reduction at (122°E, 24°N) and 71% reduction at (121°E, 22°N) when the radius was reduced from 100 to 50 km. Therefore, further studies are needed to develop an appropriate radius in consideration of the engineering impact, theoretical earthquake attenuation, historical seismic hazards, etc.

4.3 Uncertainty of the proposed approach

Developing a model to explain and simulate a complex physical problem involves observation, laboratory testing, numerical derivations and computations, explanations, etc. Due to the complexity of the physical problems, simplification and approximation are always involved to compromise between model validity and applicability. As a result, from the authors' point of view, it is nearly impossible to develop a perfect model to simulate

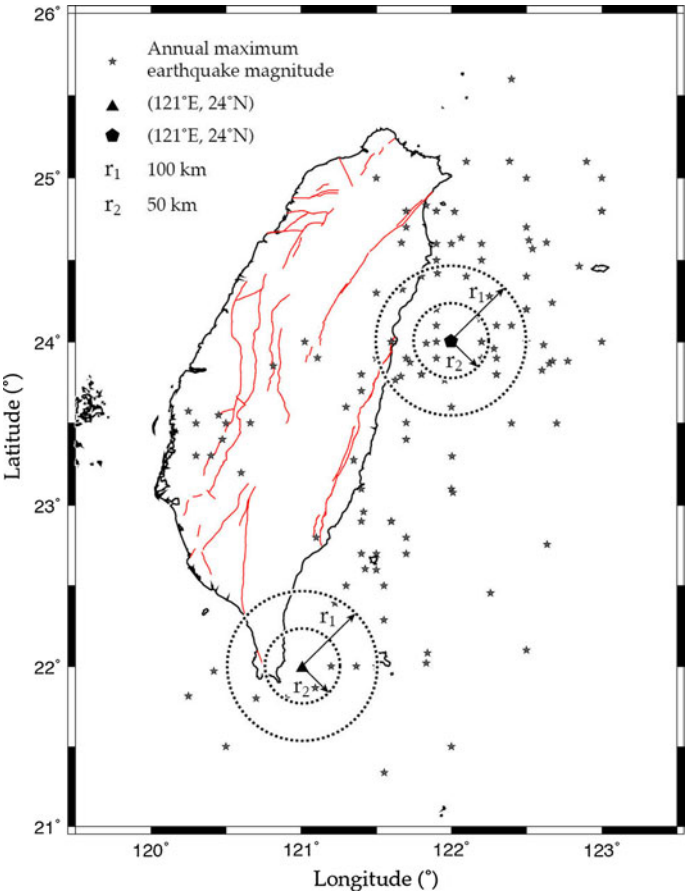

Fig. 13 Illustrations of the alternative approach to consider the uncertainty of recurrence probability (note that *the circle* was drawn for illustration purpose. Its radius was approximated to 50 and 100 km)

and forecast the natural, large-scale physical phenomena. To compensate for model uncertainty, in civil engineering for instance, a factor of safety is introduced in design. For a qualified design, the factor of safety must be greater than one, even two or three, depending on types of analysis. Theoretically speaking, the designs with factors of safety equal to 1.1 and 3 are both adequate, but, in reality, the design with the factor of safety of 1.1 is not approved by most design standards. One reason is that the model uncertainty or model imperfection is taken into account.

The uncertainty in the estimation of recurrence probability of catastrophic earthquakes when using the proposed methodology can be categorized into two components: (1) mathematical legitimacy and (2) reliability of observational data. The first component was derived and described in detail in this paper and its validity was presented and discussed. In the example provided in this study, the catalog was used "as is" without further processing. In other words, the reliability of the catalog was not studied and it would become one source of uncertainty in the estimation of recurrence probability. Therefore, to pursue a more reliable estimation, additional studies are required to verify the authenticity and reliability of the observational data, especially the observations in the early period of time

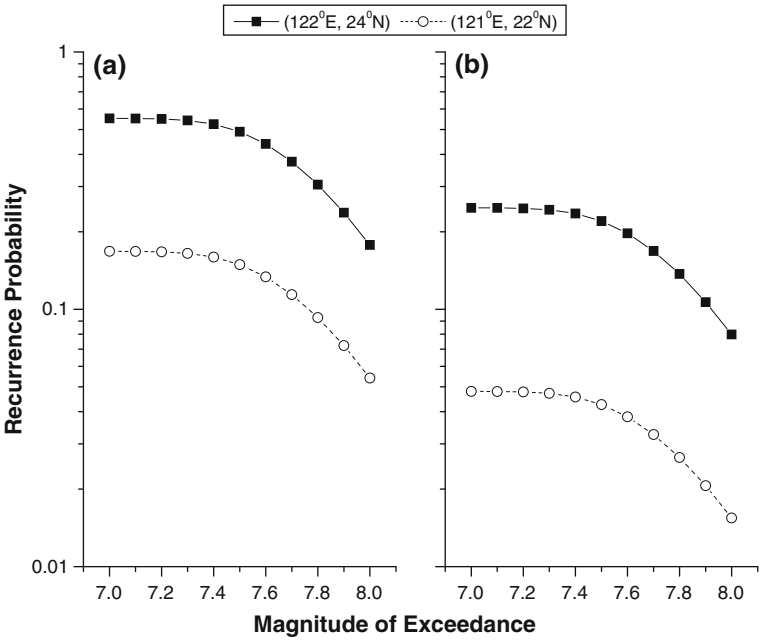

Fig. 14 Site-specific recurrence probability for two locations by using the circular probable zone in Fig. 13 with radius of 50 and 100 km

before the advent of modern instrumentation and detection techniques. Alternatively, sensitivity analysis can be performed to examine the impact of the uncertainty of early observations on the proposed approach.

5 Conclusion

In order to forecast large-size earthquake recurrence probabilities, the distribution of the annual maximum earthquake magnitude (AMEM) was examined. An approach utilizing the AMEM distribution to estimate recurrence probabilities was proposed in this paper, and an example using the earthquake catalog from 1900 to 2009 around Taiwan was illustrated. It indicates that the lognormal and Gamma distributions were found to be more suitable to model the AMEM around Taiwan. The mean value of the 110-AMEM was 6.433 associated with a coefficient of variation of 10%. Accordingly, the best-estimate recurrence probability of the earthquake with magnitude greater than 7.5 is 4% for a 1-year period. Using the proposed recurrence location model, the recurrence probability of such earthquakes with magnitude greater than 7.5 is around 5% in Central Taiwan considering a period of 50 years.

References

Anderson JG, Wesnousky SG, Stirling MW (1996) Earthquake size as a function of slip rate. Bull Seismol Soc Am 86:683–690

Ang A, Tang W (2003) Probability concept in engineering: emphasis on applications to civil and environmental engineering, 2nd edn. John Wiley & Sons, Inc., New Jersey, pp 293–294

Bommer J, Crowley H (2006) The influence of ground-motion variability in earthquake loss modeling. Bull Earthq Eng 4:231–248
Castanos H, Lomnitz C (2002) PSHA: is it science? Eng Geol 66:315–317
Cornell CA (1968) Engineering seismic risk analysis. Bull Seism Soc Am 58:1583–1606
Field EH (2007) Overview of the working group for the development of regional earthquake likelihood models (RELM). Seismol Res Lett 78(1):7–16
Jin A, Aki K (1988) Spatial and temporal correlation between Coda Q and seismicity in China. Bull Seismol Soc Am 78:741–769
Jordan T (2006) Earthquake predictability, brick by brick. Seismol Res Lett 77(1):3–6
Kijko A (2004) Estimation of the maximum earthquake magnitude, m_{max}. Pure Appl Geophys 161:1655–1681
Krintzsky EL (2002) Epistemic and aleatory uncertainty: a new shtick for probabilistic seismic hazard analysis. Eng Geol 66:157–159
Papastamatiou D (1980) Incorporation of crustal deformation to seismic hazard analysis. Bull Seismol Soc Am 70:1321–1335
Pisarenko VF, Lyubushin AA, Lysenko NB, Golubieva TV (1996) Statistical estimation of seismic hazard parameters: maximum possible magnitude and related parameters. Bull Seismol Soc Am 86:691–700
Suppe J (1984) Kinematics of arc—continent collision, flipping of subduction, and back-arc spreading near Taiwan. Mem Geol Soc China 6:21–33
U.S. Nuclear Regulatory Commission (2007) A performance-based approach to define the site-specific earthquake ground motion. NUREG-1.208, Washington
Wang CY, Shin TC (1998) Illustrating 100 years of Taiwan seismicity. Terr Atmos Oceanic Sci 9:589–614
Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture, length, rupture width, rupture area, and surface displacement. Bull Seismol Soc Am 84:974–1002
Wu YM, Chang CH, Zhao L, Teng TL, Nakamura M (2008) A Comprehensive relocation of earthquakes in Taiwan from 1991 to 2005. Bull Seism Soc Am 98:1471–1481. doi:10.1785/0120070166
Yu SB, Chen HY, Kuo LC, Lallemand SE, Tsien HH (1997) Velocity field of GPS stations in the Taiwan area. Tectonophysics 274:41–59