

Alternative representation of the Gutenberg–Richter relation in terms of the logarithmic mean annual seismicity rate and its standard deviation

Wen-Yen Chang^{1,2} · Kuei-Pao Chen² · Yi-Ben Tsai³

Received: 29 January 2016 / Accepted: 28 August 2016 / Published online: 19 December 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Gutenberg and Richter developed an empirical relation, $\log_{10} N(M) = a - bM$, to quantify the seismicity rate of various magnitudes in a given region and time period. They found the equation fit observed data well both globally and for particular regions. In conventional G–R relation, $N(M)$ represents an arithmetic mean. As a result, the arithmetic standard deviation cannot be explicitly incorporated in the log-linear G–R relation. Moreover, this representation is susceptible to influence of spuriously large numbers of aftershocks of major earthquake sequences. To overcome these shortcomings, we propose an alternative representation of the G–R relation in terms of the logarithmic mean annual seismicity rate and its standard deviation. We select the crustal earthquake data from 1973 to 2011, as listed in the National Earthquake Information Center (NEIC) global catalog and the Central Weather Bureau (CWB) Taiwan regional catalog, to illustrate our methodology. We first show that by using the logarithmic annual seismicity rates we can significantly suppress the influences of spuriously large numbers of aftershocks following major earthquake sequences contained in the Taiwan regional catalog. More significantly, both the logarithmic mean annual seismicity rate and its standard deviation can be explicitly represented in the Gutenberg–Richter relation as follows:

For global crustal seismicity: $\log_{10} N = 8.14 - 1.03M \pm (0.04M - 0.13)$;

For Taiwan crustal seismicity: $\log_{10} N = 5.62 - 0.90M \pm (0.02M + 0.17)$

where $\log_{10} N$ represents the logarithmic annual seismicity rate. Above analytical equations are very well constrained by observed global seismicity data with $5.0 \leq M \leq 7.0$ and

✉ Kuei-Pao Chen
chenkueipao@yahoo.com.tw

¹ Department of Natural Resources and Environmental Studies, NDHU, No. 1, Sec. 2, Da Hsueh Rd., Shoufeng 97401, Hualien, Taiwan, ROC

² Eastern Taiwan Earthquake Research Center (E-TEC), NDHU, No. 1, Sec. 2, Da Hsueh Rd., Shoufeng 97401, Hualien, Taiwan, ROC

³ Pacific Earthquake Science Associates, Los Altos, CA 94022, USA

by Taiwan seismicity data with $3.0 \leq M \leq 5.0$. Both equations can be extrapolated with confidence to simultaneously estimate not only the median annual seismicity rates but also their uncertainties for large earthquakes for the first time since inception of the G–R relation. These equations can be used to improve the conventional probabilistic seismic hazard assessment by including the dispersion of the annual seismicity rate. Finally, the corresponding numerical median annual seismicity rate with its upper and lower bounds obtained from above equations for $5.0 \leq M \leq 9.0$ is listed in Table 1.

Keywords Arithmetic mean · Arithmetic standard deviation · Logarithmic mean · Logarithmic standard deviation

Table 1 Observed and estimated median annual seismicity rate and return period with their dispersions for Taiwan and global crustal earthquakes

| Magnitude | Catalog | | | |
|--------------|--|--|---|---|
| | Taiwan catalog (CWB) Annual rate (event/year) | Taiwan catalog (CWB) Return period (year) | Global catalog (NEIC) Annual rate (event/year) | Global catalog (NEIC) Return period (year) |
| $M \geq 5.0$ | 24.55 | 0.041 | 1148.16 | 0.0009 |
| | 13.18 | 0.076 | 977.24 | 0.001 |
| | 7.08 | 0.14 | 831.76 | 0.0012 |
| $M \geq 5.5$ | 8.91 | 0.11 | 367.28 | 0.0027 |
| | 4.68 | 0.21 | 298.54 | 0.0033 |
| | 2.45 | 0.41 | 242.66 | 0.0041 |
| $M \geq 6.0$ | 3.24 | 0.31 | 117.49 | 0.0085 |
| | 1.66 | 0.60 | 91.20 | 0.011 |
| | 0.85 | 1.18 | 70.79 | 0.014 |
| $M \geq 6.5$ | 1.17 | 0.85 | 37.58 | 0.027 |
| | 0.59 | 1.69 | 27.86 | 0.036 |
| | 0.30 | 3.33 | 20.65 | 0.048 |
| $M \geq 7.0$ | 0.43 | 2.33 | 12.02 | 0.083 |
| | 0.21 | 4.76 | 8.51 | 0.12 |
| | 0.10 | 10.0 | 6.03 | 0.17 |
| $M \geq 7.5$ | 0.15 | 6.67 | 3.85 | 0.26 |
| | 0.074 | 13.51 | 2.60 | 0.38 |
| | 0.036 | 27.78 | 1.76 | 0.57 |
| $M \geq 8.0$ | 0.056 | 17.86 | 1.23 | 0.81 |
| | 0.026 | 38.46 | 0.79 | 1.27 |
| | 0.012 | 83.33 | 0.51 | 1.96 |
| $M \geq 8.5$ | 0.020 | 50.00 | 0.39 | 2.56 |
| | 0.009 | 111.11 | 0.24 | 4.17 |
| | 0.004 | 250.0 | 0.15 | 6.67 |
| $M \geq 9.0$ | 0.0074 | 135.14 | 0.13 | 7.69 |
| | 0.0033 | 303.03 | 0.074 | 13.51 |
| | 0.0015 | 666.67 | 0.04 | 25.00 |

Observed value is shown in bold number, estimated value in regular number

$\log_{10} N = 5.62 - 0.90M \pm (0.02M + 0.17)$ for Taiwan crustal earthquakes

$\log_{10} N = 8.14 - 1.03M \pm (0.04M - 0.13)$ for global crustal earthquakes

1 Introduction

Gutenberg and Richter (1941, 1944) developed an empirical relation, $\log_{10} N(M) = a - bM$, to quantify the seismicity rate of various magnitudes in any given region and time period, where $N(M)$ represents the cumulative number of earthquakes per year with magnitudes equal or greater than M , and a and b are constants. They found the equation fit observed data well both globally and for particular regions. For example, Gutenberg and Richter (1944) applied the formula $\log_{10} N = a + b(8 - M)$ to the southern California earthquake data for the period January 1934–May 1943 by linear least-squares regression. They obtained $a = -2.04 \pm 0.09$, $b = 0.88 \pm 0.03$. In their formula, N represented the mean annual number of earthquakes per 0.1 unit of M . A similar procedure was applied later by Gutenberg and Richter (1954) to obtain the following values of a and b from global earthquakes at different depth ranges: (1) shallow shocks: $a = -0.48 \pm 0.02$, $b = 0.90 \pm 0.02$; (2) intermediate-depth shocks: $a = -1.2 \pm 0.2$, $b = 1.2 \pm 0.2$; (3) deep shocks: $a = -1.9 \pm 0.2$, $b = 1.2 \pm 0.2$. It should be pointed out that these a and b values were all referred to the arithmetic mean annual seismicity rate. In addition to the linear least-squares regression, another commonly used method developed by Aki (1965) is to make maximum likelihood estimate of the b value in the G–R relation and its confidence limits.

Since then the Gutenberg–Richter (G–R) relation has been widely used in quantitative seismicity studies. Here are several recent examples: Hutton et al. (2010) used the G–R relation to determine the magnitude of completeness (M_c) of the southern California earthquake catalog over different periods from 1932 to 2008. Michael (2014) recently used it to check the completeness of the ISC–GEM global earthquake catalog. Konsuk and Aktas (2013) also made use of the G–R relation to estimate the recurrence period of earthquakes in Turkey. The G–R relation also plays a major role in probabilistic seismic hazard analysis (Reiter 1990). For example, Frankel et al. (1996) included the G–R relation in their original development and subsequent updates of the US National Seismic Hazard Maps.

It should be pointed out that in previous studies the a and b values in the G–R relation were all determined by applying either the least-squares regression or the maximum likelihood method on the mean annual seismicity rate as a function of magnitude. The annual seismicity rate in the conventional G–R relation represents an arithmetic mean (AM), as it is commonly obtained by dividing the total number of earthquakes with the number of years covered in a catalog. It does not account for the variability in individual annual seismicity rates. In this case, the arithmetic standard deviation is incompatible with the log-linear G–R relation and thus cannot be explicitly incorporated therein. Moreover, it has been found that the AM representation is susceptible to significant influence of spuriously large numbers of aftershocks of major earthquake sequences.

In order to avoid these major shortcomings, we propose an alternative representation of the G–R relation in terms of the logarithmic mean annual seismicity rate and its standard deviation. Instead of obtaining the overall arithmetic mean annual seismicity rate for $N(M)$, we begin by plotting in series the logarithmic annual seismicity rate of individual year, $\log_{10} N(M)$, chronologically throughout the whole period of the catalog for earthquakes with magnitudes equal or greater than a given M . We then calculate the mean and standard deviation of this series. By plotting these logarithmic mean and standard deviation values for a series of M on a $\log_{10} N(M)$ versus M chart, we can obtain the G–R relations for the logarithmic mean annual seismicity rate and its \pm standard deviation, respectively. These three equations are finally combined into a single G–R relation expressed explicitly in terms of the logarithmic mean annual seismicity rate and its standard deviation.

In order to illustrate our simple methodology and its merits, we select two contemporary instrumental earthquake catalogs from 1973 to 2011 to cover crustal earthquakes in different magnitude ranges over very different size of areas. First, the CWB catalog was compiled by combining the seismicity data obtained by the Taiwan Telemetered Seismic Network (1973–1991) and the Central Weather Bureau Seismic Network (1992–2011) in Taiwan region. The original M_D and M_L magnitudes in this catalog have been further converted to homogeneous M_w (Chen and Tsai 2008; Chang et al. 2016). Another catalog was compiled by the

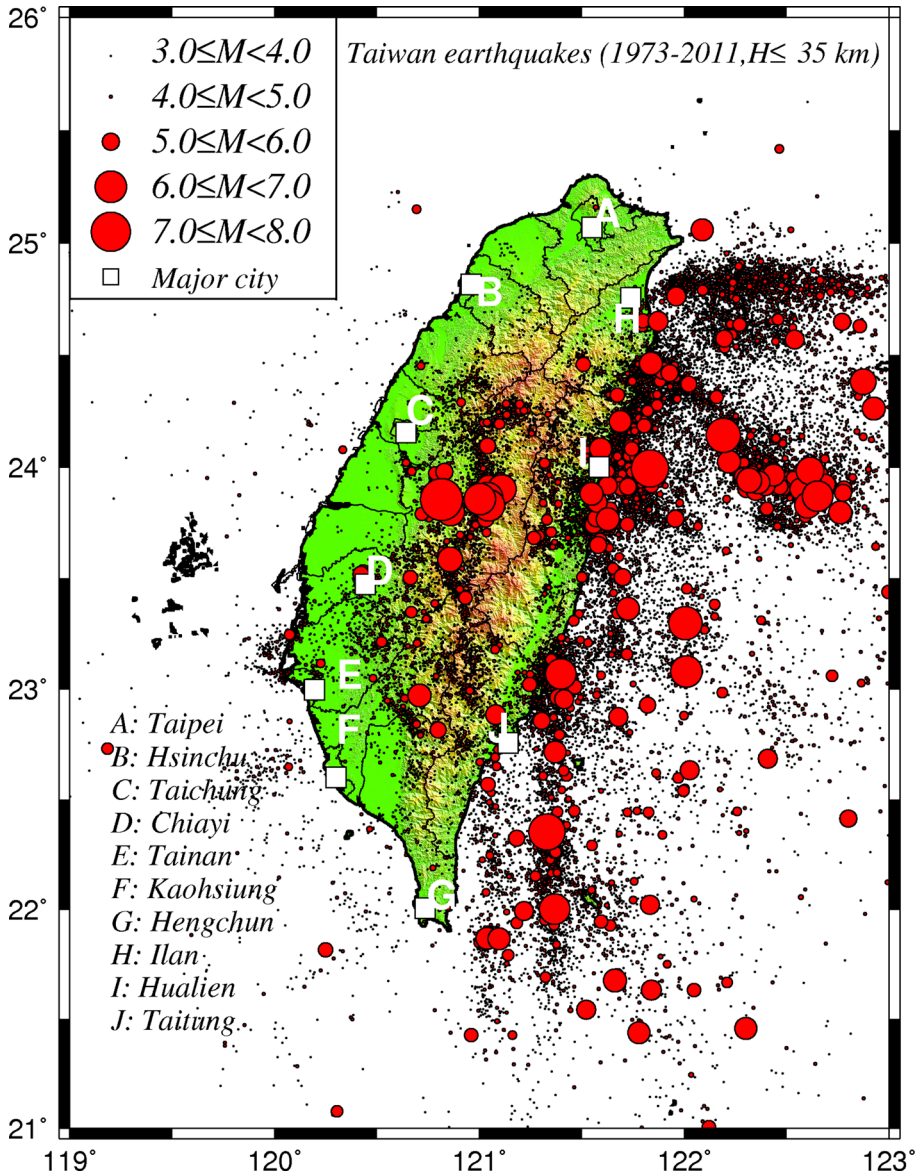


Fig. 1 Epicenter distribution of Taiwan crustal earthquakes with magnitude $M \geq 3.0$ from 1973 to 2011 (after Chen and Tsai 2008; Chang et al. 2016)

National Earthquake Information Center (NEIC) which used actual or adopted M_w for global earthquakes. These two earthquake catalogs are arguably among the most complete and homogeneous catalogs available in the world. In the followings, we use M to represent the moment magnitude, as both the CWB and NEIC catalogs contain some actual (Harvard CMT), adopted (NEIC) or converted (CWB) moment magnitude values.

In the followings, we will first begin by comparing the logarithmic mean annual seismicity rate and its standard deviation with their arithmetic counterparts, all of which are obtained directly from series of the annual seismicity rate in individual years of different magnitude ranges from both catalogs. Next, we proceed to show that the annual seismicity rates can be fitted better with a lognormal distribution than a normal one, especially for the Taiwan regional data set, which contains spuriously large numbers of aftershocks from two major earthquake sequences. Finally, we will obtain the G–R relations expressed in terms of the logarithmic mean and its standard deviation for crustal earthquakes from these two regional and global catalogs, respectively. At the end, we will present a numerical table listing the corresponding observed and estimated median annual seismicity rates and its upper and lower bounds at \pm one standard deviation, as calculated from these alternative G–R relations for $5.0 \leq M \leq 9.0$.

2 Comparison between the logarithmic mean annual seismicity rate and its standard deviation with their arithmetic counterparts

We begin by comparing the logarithmic mean annual seismicity rate and its standard deviation with their arithmetic counterparts, all of which are obtained directly from series of the annual seismicity rate in individual year for different magnitude ranges. We use two

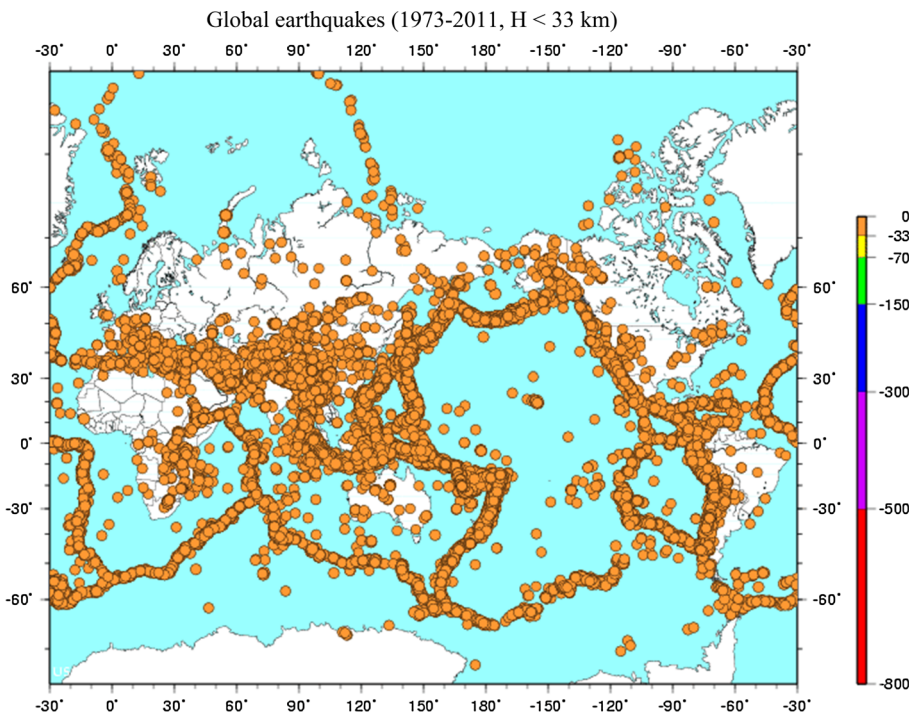


Fig. 2 Epicenter distribution of global crustal earthquakes with magnitude $M \geq 5.0$ from 1973 to 2011 (from NEIC Web site: <http://earthquake.usgs.gov/earthquakes/search/>)

Fig. 3 **a** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 3.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **b** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 3.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation. **c** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 4.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **d** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 4.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation. **e** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **f** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation. **g** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **h** Annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation

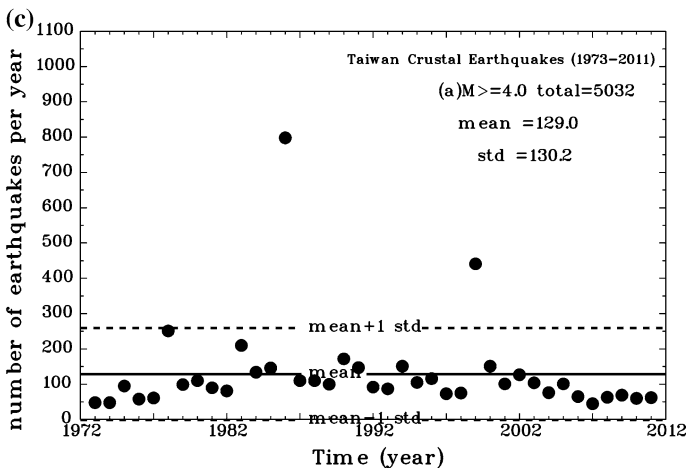
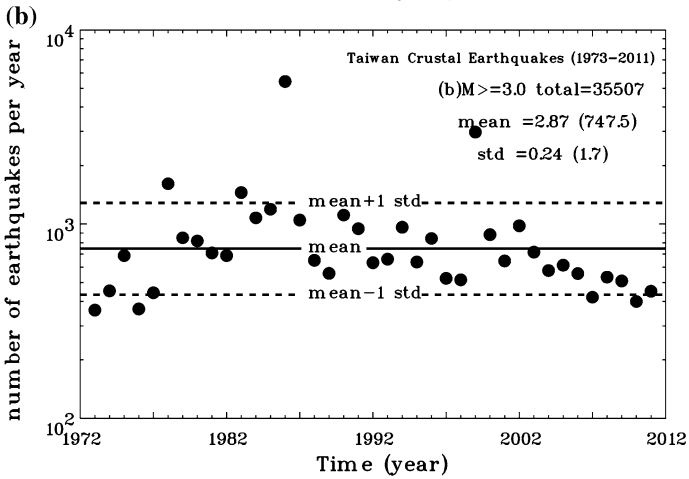
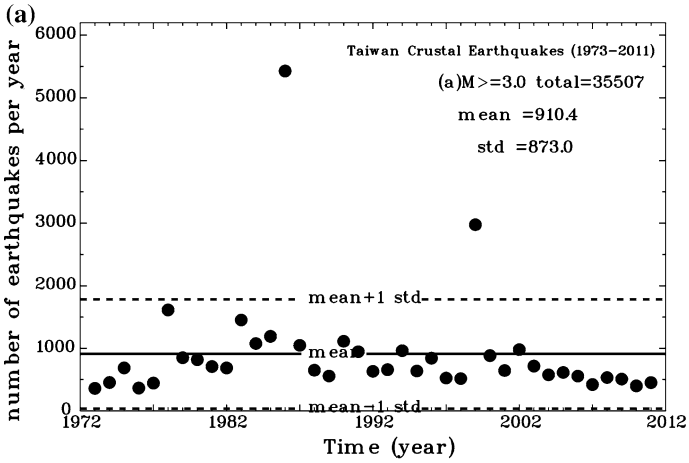
crustal earthquake data sets based on the CWB Taiwan earthquake catalog with homogenized moment magnitudes (Chen and Tsai 2008; Chang et al. 2016) and the NEIC global earthquake catalog. Both data sets cover the same time period from 1973 to 2011, as shown in Figs. 1 and 2, respectively. The former data set provides sufficient seismicity data from $3.0 \leq M \leq 5.0$, whereas the latter one covers the magnitude range from $5.0 \leq M \leq 7.0$, for our methodology to make meaningful comparisons.

First, we use the logarithmic and arithmetic earthquake numbers in individual years to calculate the corresponding mean annual seismicity rate and its standard deviation for Taiwan crustal earthquakes, as shown in Fig. 3a–h. Figure 3a, b shows a total of 35,507 earthquakes with $M \geq 3.0$ that had occurred in Taiwan region from 1973 to 2011. During this period, there were two major earthquake sequences in 1986 and 1999, which had produced over 5500 and 3000 aftershocks, respectively. As a result, the arithmetic mean annual seismicity rate of 910.4 is apparently overestimated and biased above the main body of the data population, as shown in Fig. 3a. In the meantime, its standard deviation of 873.0 is also significantly increased, so that only the two largest data points lie far above the mean-plus-one-standard-deviation level. On the other hand, Fig. 3b shows that both the logarithmic mean of 2.87 (with a corresponding median value of 747.5 event/year) and its standard deviation of 0.24 (with a corresponding multiplication factor of 1.7) match much better with bulk of the data population. Similar observations can be said by comparing Fig. 3c, d for $M \geq 4.0$ and Fig. 3e, f for $M \geq 5.0$ respectively.

Finally, Fig. 3g, h shows the series of annual seismicity rate for $M \geq 6.0$ as plotted on linear and logarithmic scales, respectively. We can see zero count or just one count of events in many years over the series, resulting in the arithmetic-mean-minus-one-standard-deviation value to become negative. In this case, the logarithmic measure cannot be applied or will yield many zero values. These examples suggest that the mean annual seismicity rate needs to be greater than about 10 for our method to produce robust results.

In summary, from above comparisons we can see the logarithmic mean value is smaller than its arithmetic counterpart, as expected by the AM-GM (geometric mean) inequality. More significantly, the logarithmic standard deviation is much smaller than its arithmetic counterpart for the magnitude range $3.0 \leq M \leq 5.0$ where sufficient data are available for logarithmic measure. Both the arithmetic mean and its standard deviation are significantly influenced by spuriously large number of aftershocks from the two major earthquake sequences in 1986 and 1999, respectively. On the other hand, the logarithmic measures are able to significantly suppress the influences by spuriously large numbers of aftershocks in individual years.

Next we use the NEIC global crustal earthquake data for similar comparisons over a range of higher magnitudes and a much larger area of coverage. Figure 4a–h shows series



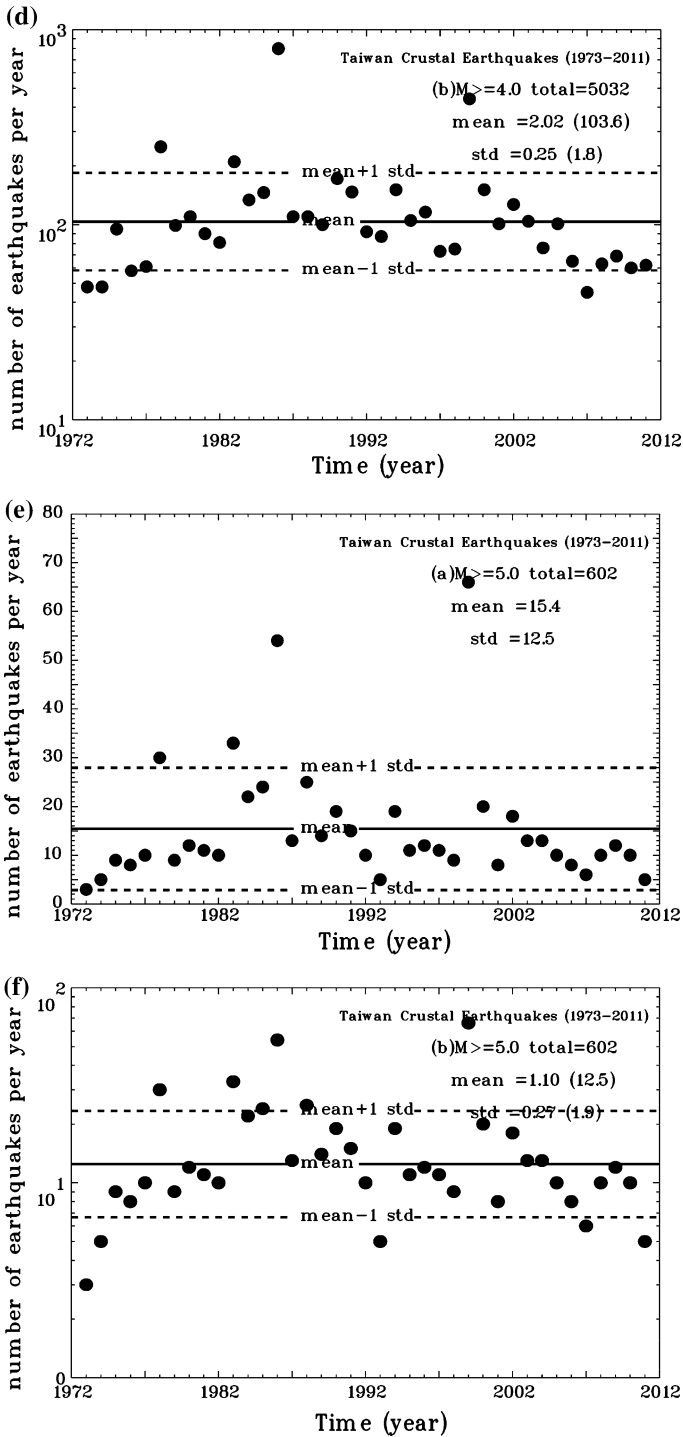


Fig. 3 continued

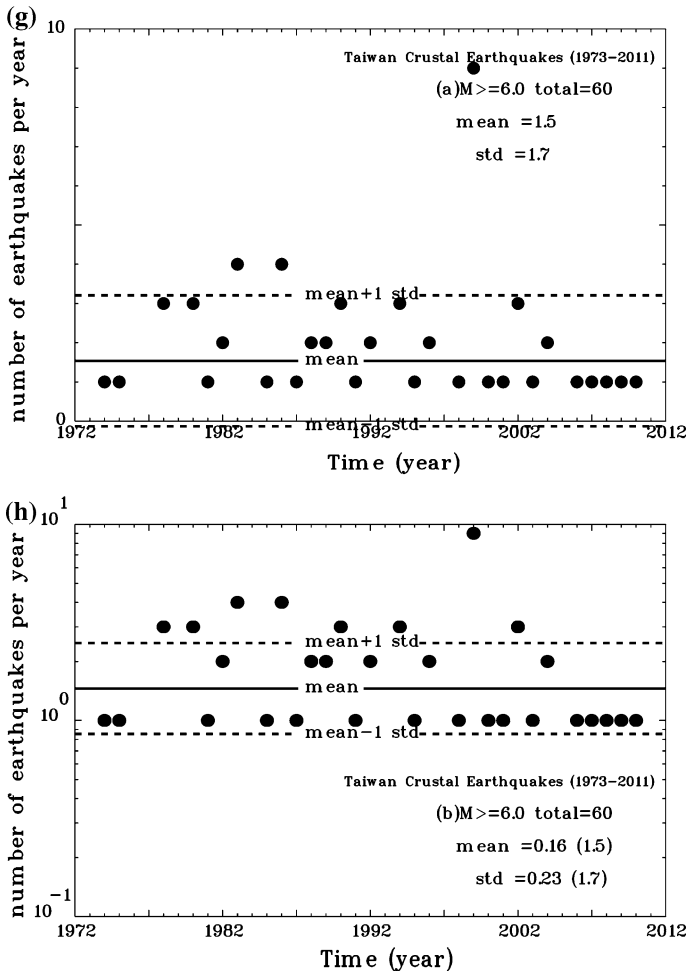


Fig. 3 continued

of annual seismicity rate from 1973 to 2011 for global crustal earthquakes with $5.0 \leq M \leq 8.0$, as plotted on linear and logarithmic scales, respectively. Figure 4a, b shows that a total of 36,089 crustal earthquakes with $M \geq 5.0$ took place globally from 1973 to 2011, resulting in an arithmetic mean of 925.4 with a \pm one standard deviation range of 337.4 (from 756.7 to 1094.1), and a logarithmic mean of 2.96 with a standard deviation of 0.08, equivalent to a multiplication factor of 1.2. The logarithmic values give a corresponding median of 910.8 with a \pm one standard deviation range of 334.0 (from 759.0 to 1093.0). In this case, the logarithmic mean annual seismicity rate and its standard deviation yield slightly smaller values than their arithmetic counterparts, consistent with the AM–GM inequality. Similar observations can be said for the cases of $M \geq 6.0$ in Fig. 4c, d and of $M \geq 7.0$ in Fig. 4e, f respectively.

Finally, Fig. 4g, h shows the series of annual seismicity rate for earthquakes with $M \geq 8.0$ as plotted on linear and logarithmic scale, respectively. Figure 4g shows zero count or just one count of event occurrence in many individual years, resulting in the

Fig. 4 **a** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **b** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation. **c** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **d** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation. **e** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 7.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **f** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 7.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation. **g** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 8.0$ from 1973 to 2011 with corresponding arithmetic mean and standard deviation. **h** Annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 8.0$ from 1973 to 2011 with corresponding logarithmic mean and standard deviation

arithmetic-mean-minus-standard-deviation value to become negative. For these years, the logarithmic measure cannot be applied or will yield zero values. Like the previous case of Taiwan seismicity data, our methodology can yield robust determination of the logarithmic mean annual seismicity rate and its standard deviation for a magnitude range only when sufficiently large annual seismicity rates in individual years, say more than 10 events per year, are available.

In summary, we can see from above comparisons the logarithmic mean and standard deviation can yield slightly better measurement of the annual seismicity rate than their arithmetic counterparts for global crustal earthquakes in the magnitude range of $5.0 \leq M \leq 7.0$. For $M \geq 8.0$, the annual event numbers become too sparse for our method to be applicable.

3 Lognormal versus normal distributions of the annual seismicity rates

Next we proceed to compare whether the annual seismicity rates shown previously in Figs. 3 and 4 fit better with a lognormal or a normal distribution function. For this purpose, we apply least-squares fitting to the observed ensembles of annual seismicity rate, as follows:

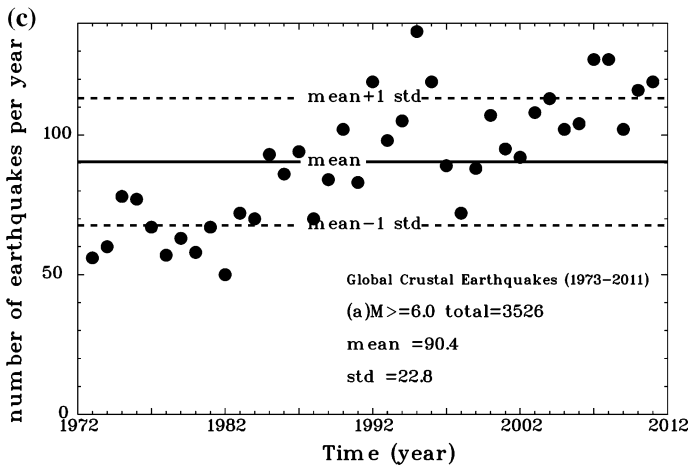
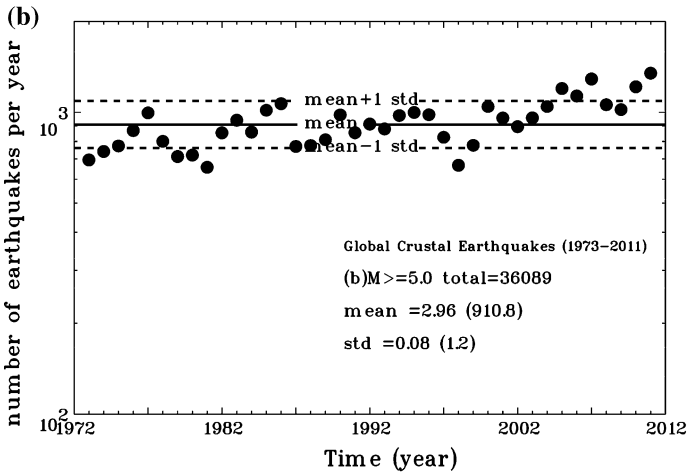
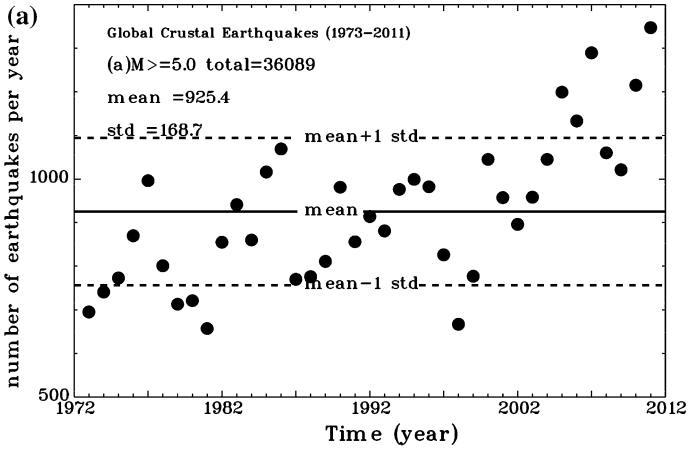
$$Y(x_i) = ay(x_i) \quad (1)$$

$$a = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N y(x_i)} \quad (2)$$

where Y_i = observed values, $y(x_i) = e^{-\frac{(x-\mu)^2}{2\delta^2}}$, μ = mean and δ = standard deviation, as given in Figs. 3 and 4. Then, we use the calculated a value to replace $\frac{1}{\sqrt{2\pi\delta^2}}$ and

$R^2 = \sum_{i=1}^N [Y_i - ay(x_i)]^2$, where R is the root-mean-square (RMS) error.

We aggregate the annual seismicity rates of Taiwan crustal earthquakes with $M \geq 3.0$, $M \geq 4.0$, $M \geq 5.0$ and $M \geq 6.0$, in equal bins of 100, 10, 2, 1, respectively, for the normal distribution, and an equal bin of 0.05 for the lognormal distribution. The results are shown in Fig. 5a–h. Figure 5a clearly shows the peak of the normal distribution curve is offset to the right of the bulk of data population, primarily due to two large annual seismicity rates in 1986 and 1999, respectively. This results in a relatively large RMS error of 0.77. In contrary, Fig. 5b shows the peak of the lognormal distribution curve is centered in the bulk of data population of $M \geq 3.0$, with a much reduced RMS error of 0.61. Similar



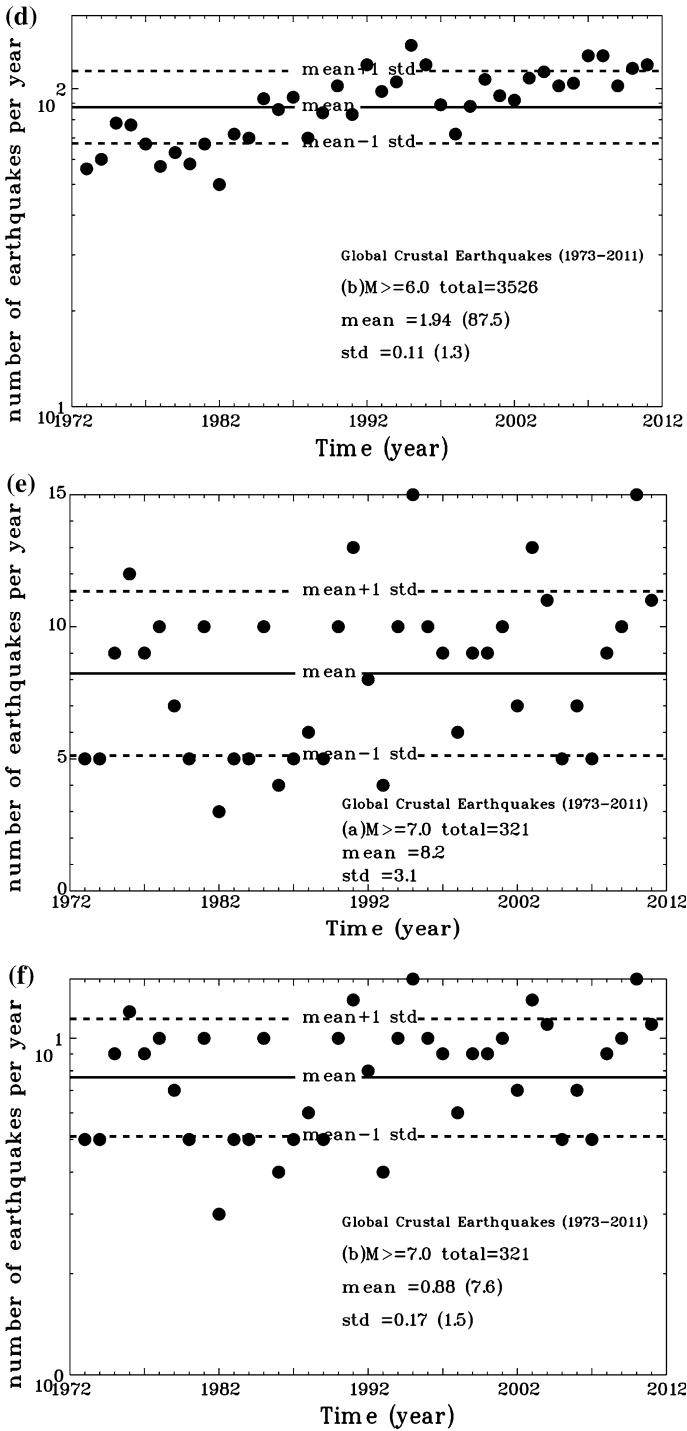


Fig. 4 continued

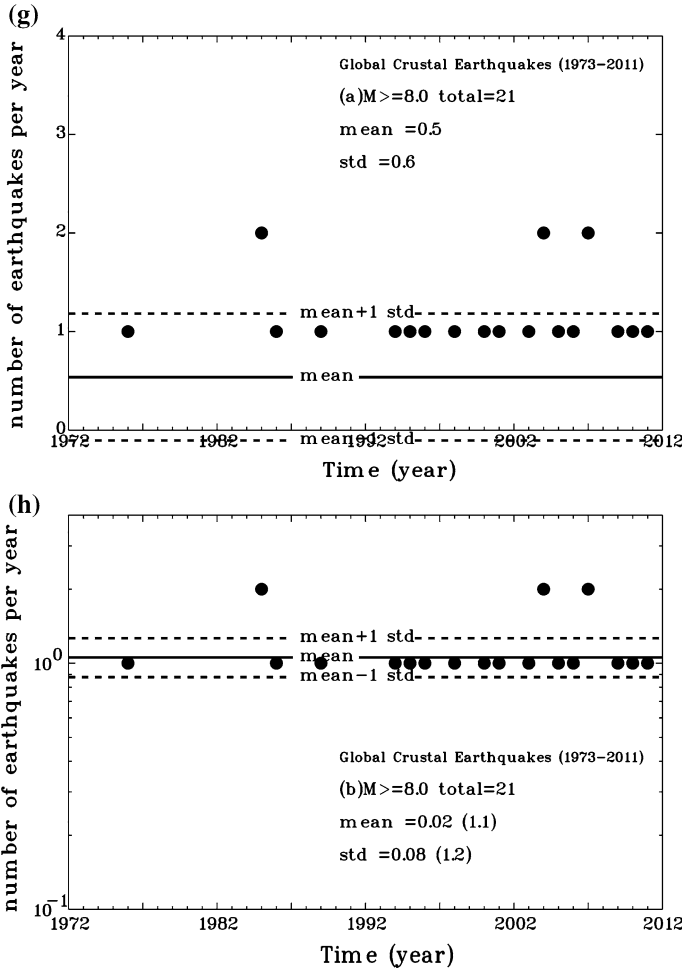


Fig. 4 continued

observations can be said about the matching between the observed data populations and their respective probability distribution curves, as shown in Fig. 5c, d for $M \geq 4.0$ and in Fig. 5e, f for $M \geq 5.0$, respectively. Finally, Fig. 5g, h shows breakdown of such matchings for $M \geq 6.0$ because available observed data are too sparse.

In summary, we can see from above comparisons the observed annual seismicity rates for Taiwan crustal earthquakes can be fitted significantly better with a lognormal distribution than a normal one, as judged by the RMS errors. We can also see from Fig. 5g, h such matching would breakdown when available observed data are too sparse.

Similarly, we apply the same process to the global crustal earthquake data set. We select data from $5.0 \leq M \leq 8.0$ on the basis of their completeness. In order to fit these observed data with a probability density function, we aggregate them in equal bins of 100, 10, 1, 1 for $M \geq 5.0$, $M \geq 6.0$, $M \geq 7.0$ and $M \geq 8.0$, respectively, for the normal distribution, and in an equal bin of 0.05 for the lognormal distribution. The results are shown in Fig. 6a–h. We can see from Fig. 6a–f the RMS errors are small and comparable to each other between

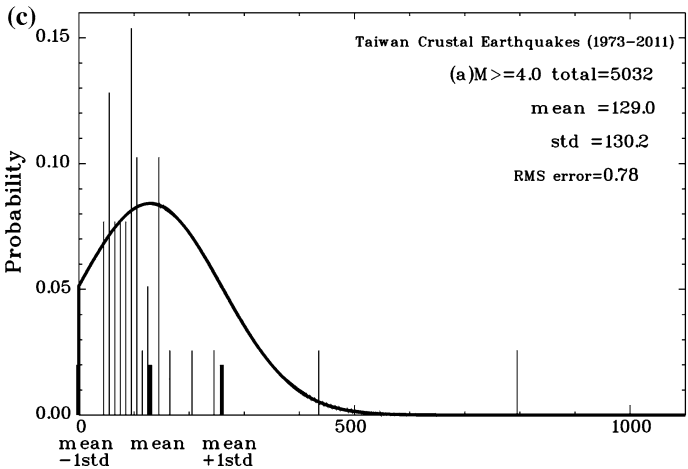
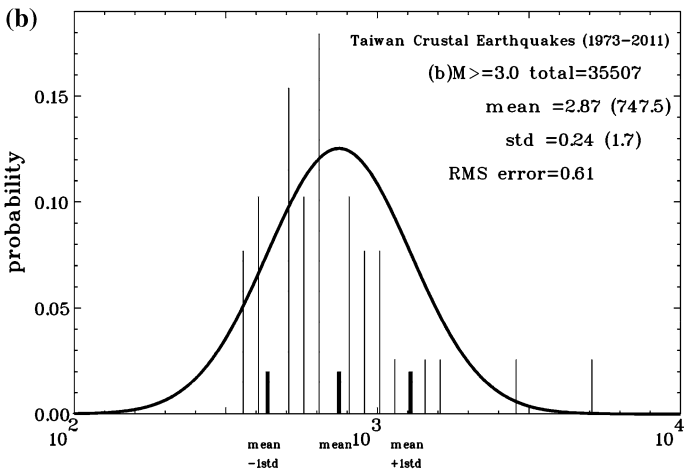
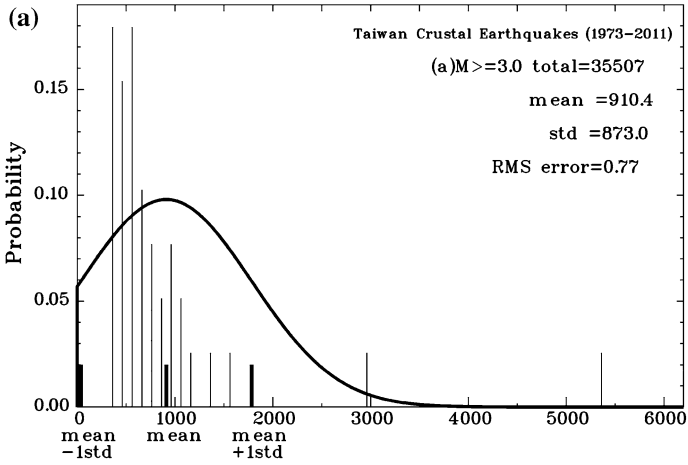
Fig. 5 **a** Least-squares fitting with normal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 3.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3a. **b** Least-squares fitting with lognormal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 3.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3b. **c** Least-squares fitting with normal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 4.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3c. **d** Least-squares fitting with lognormal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 4.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3d. **e** Least-squares fitting with normal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3e. **f** Least-squares fitting with lognormal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3f. **g** Least-squares fitting with normal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3g. **h** Least-squares fitting with lognormal distributions of the annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 3h

the lognormal and normal distributions for the three magnitude ranges of $M \geq 5.0$, $M \geq 6.0$ and $M \geq 7.0$. This is largely due to the absence of spuriously large annual seismicity rates in individual years. Figure 6g, h again shows the available observed data for $M \geq 8.0$ are too sparse to allow for meaningful probability assessment.

4 Alternative representation of the Gutenberg–Richter relation in terms of the logarithmic mean annual seismicity rate and its standard deviation

Realistic estimation of the annual seismicity rate of large future earthquakes is an important issue in probabilistic seismic hazard analysis. The Gutenberg–Richter (G–R) relation based on the arithmetic mean is commonly used for this purpose. However, the mean annual seismicity rate of large future earthquakes can be overestimated with large dispersion by using this conventional method, as shown above with the data set of Taiwan crustal earthquakes, when the catalog contains spuriously large number of aftershocks from major earthquake sequences. In order to reduce this undesirable influence, various methods have been proposed to purge aftershock data from earthquake catalogs (i.e., Reasenberg 1985; Reasenberg and Jones 1989). In this study, we propose an alternative approach which can significantly suppress the influence of large number of aftershocks by analytically representing the G–R relation in terms of the logarithmic mean annual seismicity rate and its standard deviation. More significantly, this analytical representation can allow us for the first time to estimate not only the median annual seismicity rate but also its dispersion for any given magnitudes.

Once again, we use the CWB Taiwan regional earthquake catalog and the NEIC global earthquake catalog from 1973 to 2011 to illustrate the methodology and merits of our approach. For Taiwan crustal earthquakes, we select the data with $M \geq 3.0$ to calculate both mean annual seismicity rates and their standard deviations directly from series of annual seismicity rates in individual years at a magnitude increment of $\Delta M = 0.1$. The results are plotted in the format of Gutenberg–Richter relation in Fig. 7. From the figure, we can see the arithmetic means are greater than its logarithmic counterparts. More seriously, the arithmetic standard deviation is not only large but also not symmetric about its mean. This latter feature makes it impossible for us to incorporate the arithmetic standard deviation explicitly in the G–R relation.



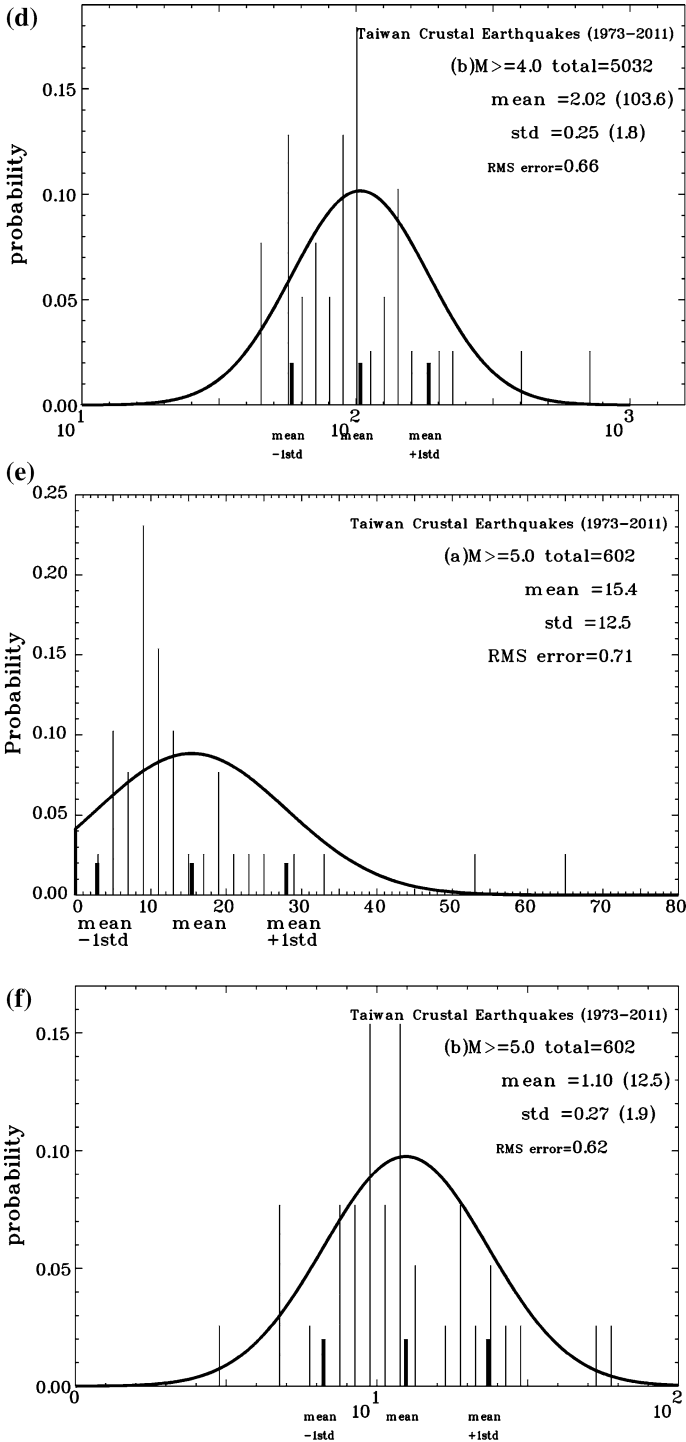


Fig. 5 continued

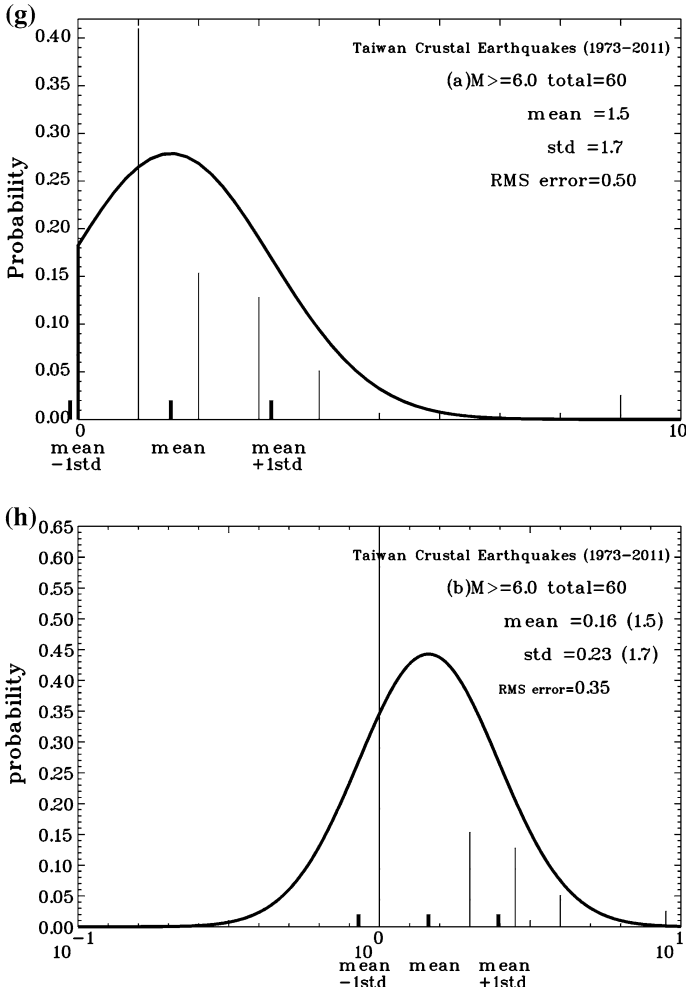


Fig. 5 continued

On the other hand, we can see in the same figure both the logarithmic mean annual rate and its standard deviation are smaller than their arithmetic counterparts. More significantly, the logarithmic standard deviation is symmetric about the logarithmic mean, especially in the magnitude range from $3.0 \leq M \leq 5.0$ where sufficient observed data are available. In this case, the logarithmic standard deviation can be incorporated explicitly in the Gutenberg–Richter relation. We can obtain from the observed data by regression from $M \geq 3.0$ to $M \geq 5.0$ the following G–R relations:

$$\text{Logarithmic mean + std. } \log_{10} N = 5.79 - 0.88M \tag{3}$$

$$\text{Logarithmic mean } \log_{10} N = 5.62 - 0.90M \tag{4}$$

Fig. 6 **a** Least-squares fitting with normal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4a. **b** Least-squares fitting with lognormal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 5.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4b. **c** Least-squares fitting with normal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4c. **d** Least-squares fitting with lognormal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 6.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4d. **e** Least-squares fitting with normal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 7.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4e. **f** Least-squares fitting with lognormal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 7.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4f. **g** Least-squares fitting with normal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 8.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4g. **h** Least-squares fitting with lognormal distributions of the annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 8.0$ from 1973 to 2011. The mean and standard deviation are the same as Fig. 4h

$$\text{Logarithmic mean} - \text{std. } \log_{10} N = 5.45 - 0.93M \quad (5)$$

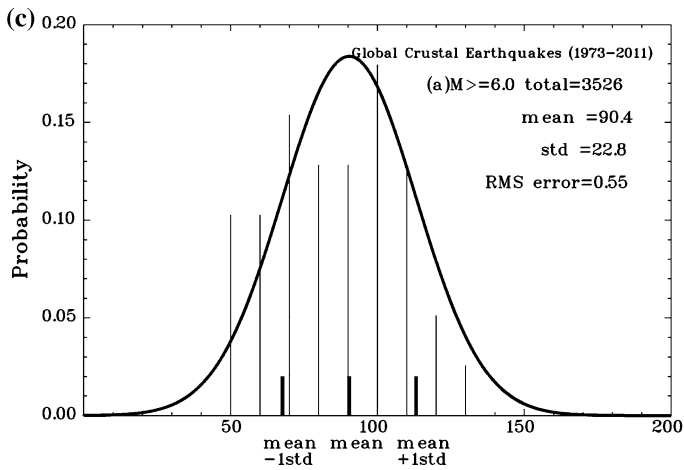
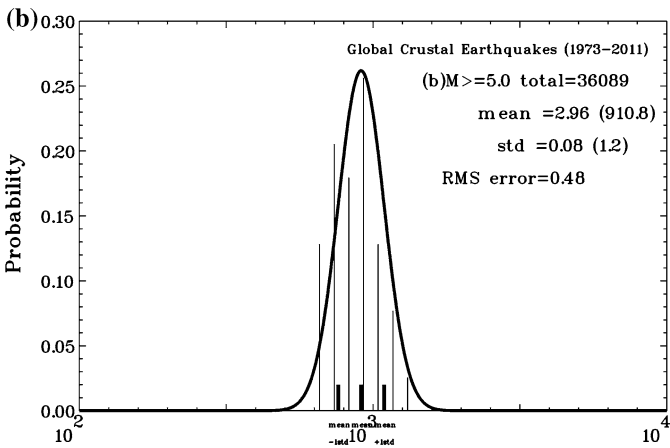
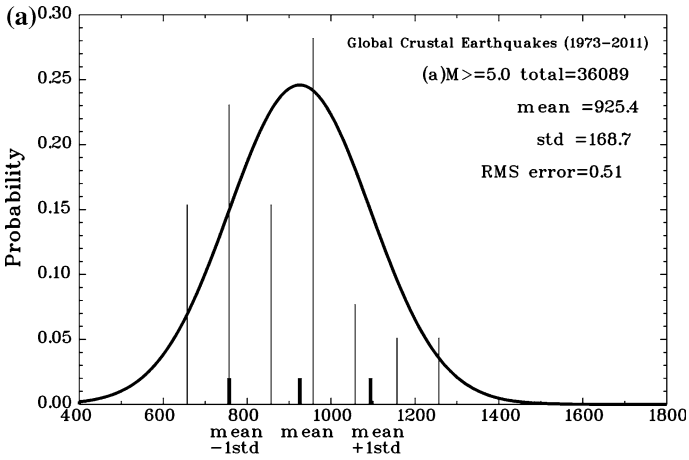
$$\text{Combined: } (3-5) \log_{10} N = 5.62 - 0.90M \pm (0.02M + 0.17) \quad (6)$$

where $\log_{10} N$ represents the logarithmic annual seismicity rate.

Equations 3 to 5 are plotted in Fig. 7, in solid lines constrained by the observed data and in dashed lines by extrapolation. These three equations can be combined into one single Eq. 6 owing to the symmetry of Eqs. 3 and 5 with respect to Eq. 4. Judging from the small standard deviations and a nearly unity of R^2 values from least-squares regression, above equations are very well constrained by the observed data from $M \geq 3.0$ to $M \geq 5.0$, resulting in robust determination of both a and b values. Accordingly, the analytical Eq. 6 can be extrapolated with confidence to obtain for the first time robust estimates of not only the logarithmic mean annual seismicity rate but also its dispersion for Taiwan crustal earthquakes with $M \geq 5.0$, as shown in Table 1. This is further confirmed below by the NEIC global crustal earthquake data set, which provides sufficient data from $5.0 \leq M \leq 7.0$ to constrain the alternative Gutenberg–Richter relation.

The observed conventional arithmetic means are also shown for comparison in Fig. 7 in open circles with their corresponding regression line. It is noticed from the figure that the observed conventional arithmetic means are enveloped by the logarithmic mean \pm standard deviation lines. The portion from $M3.0$ to $M6.0$ lies above and $M6.0$ to $M7.6$ below the logarithmic mean line, respectively. The deficiency of observed events in larger magnitude range is probably due to the relative short period of coverage by the CWB catalog we selected for this study, so that not enough numbers of large earthquakes are included. Accordingly, the estimations obtained by Eq. 6 will encompass the observed arithmetic mean annual seismicity rates from $M3.0$ to $M7.6$ for Taiwan earthquakes. It should be pointed out further that Eq. 6 shows an increasing logarithmic standard deviation with magnitude. This would result in greater dispersion in the estimation of annual seismicity rates for large Taiwan crustal earthquakes.

For global seismicity, we select from the NEIC catalog the crustal earthquakes with focal depths in the 0–33 km range and $M \geq 5.0$ for its completeness. Both the logarithmic and arithmetic means and their standard deviations at an increment of $\Delta M = 0.1$ are plotted in the format of Gutenberg–Richter relation in Fig. 8. From the figure, we can see both the logarithmic and arithmetic means and standard deviations differ only slightly from each other for earthquakes with magnitudes $5.0 \leq M \leq 7.0$ where sufficient observed data



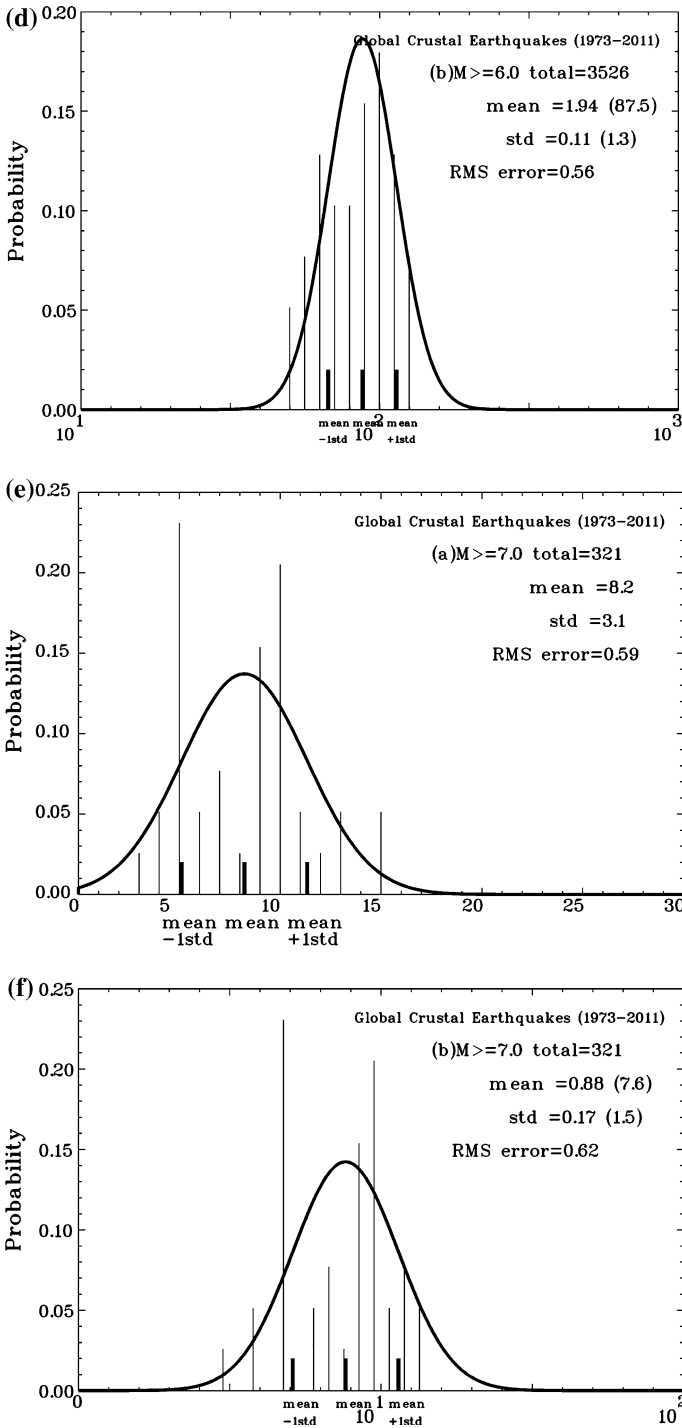


Fig. 6 continued

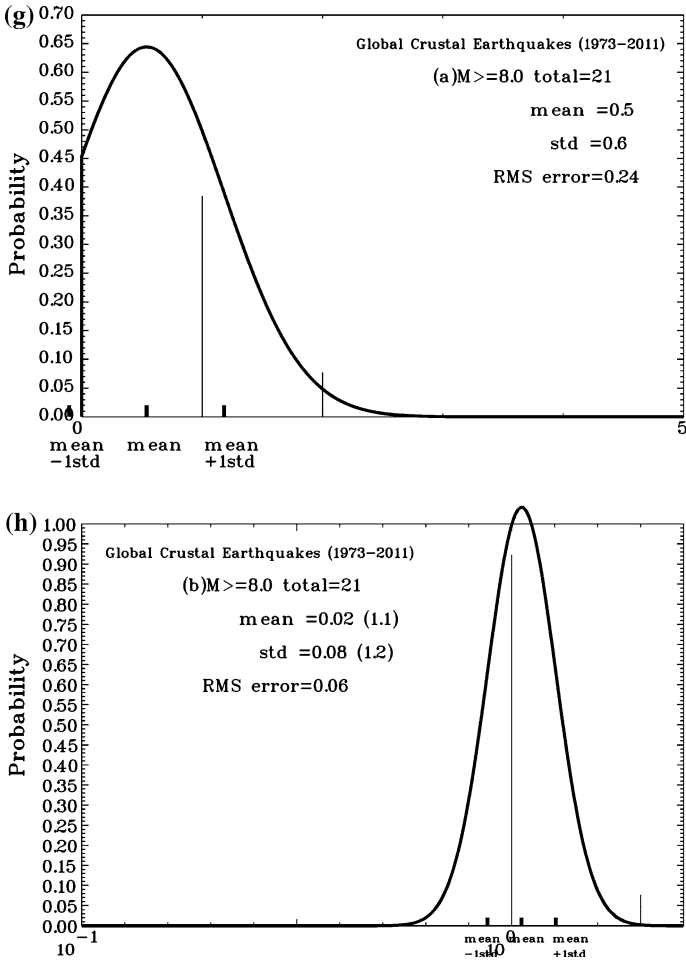


Fig. 6 continued

are available. Again only the logarithmic standard deviation can be incorporated explicitly in the Gutenberg–Richter relation owing to its symmetry with respect to the logarithmic mean. We can obtain from the observed data by regression from $5.0 \leq M \leq 7.0$ the following G–R relations at different logarithmic levels:

$$\text{Mean} + \text{std. } \log_{10} N = 8.01 - 0.99M \tag{7}$$

$$\text{Mean } \log_{10} N = 8.14 - 1.03M \tag{8}$$

$$\text{Mean} - \text{std. } \log_{10} N = 8.28 - 1.08M \tag{9}$$

$$\text{Combined: } (7-9) \log_{10} N = 8.14 - 1.03M \pm (0.04M - 0.13) \tag{10}$$

where $\log_{10} N$ represents the logarithmic annual seismicity rate.

Equations 7 to 9 are plotted in Fig. 8, in solid line constrained by observed data from $M5.0$ to $M7.0$ and in dashed line by extrapolation from $M7.0$ to $M9.0$. These three

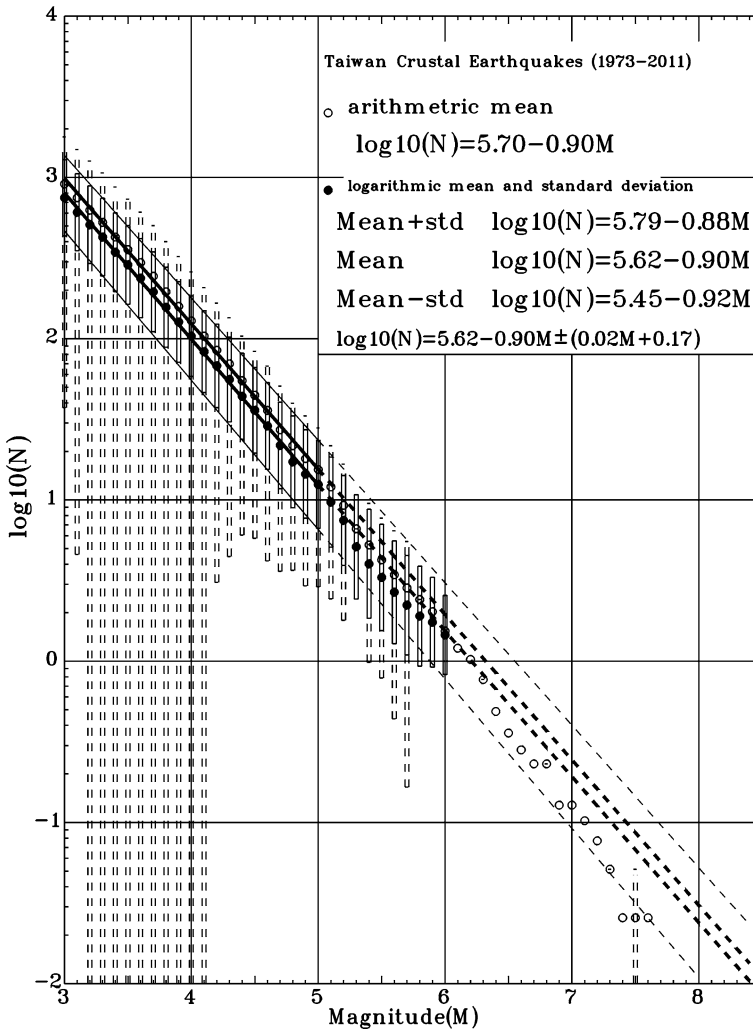


Fig. 7 Plots of the logarithmic and arithmetic mean annual seismicity rates of Taiwan crustal earthquakes with magnitudes $M \geq 3.0$ data. *Open circle* represents the observed arithmetic mean, and *black dot* represents the observed logarithmic mean. The *solid bar* represents the *upper* and *lower* bounds of logarithmic mean, and the *dashed bar* represents the *upper* and *lower* bounds of arithmetic mean both at \pm one standard deviation. The standard deviation of logarithmic mean is not only symmetric with respect to the mean, but also smaller than the standard deviation of arithmetic mean. The *solid lines* represent regression constrained by observed data, whereas *dashed lines* represent extrapolations

equations can further be combined into one single Eq. 10 owing to the symmetry of Eqs. 7 and 9 with respect to Eq. 8. Judging from the small standard deviations and near-unity R^2 values from least-squares regression, above equations are very well constrained by the observed data from $5.0 \leq M \leq 7.0$, resulting in robust determination of both a and b values. Accordingly, the analytical Eq. 10 can be extrapolated with confidence to give estimates for the first time since inception of the G–R relation, not only the logarithmic mean annual

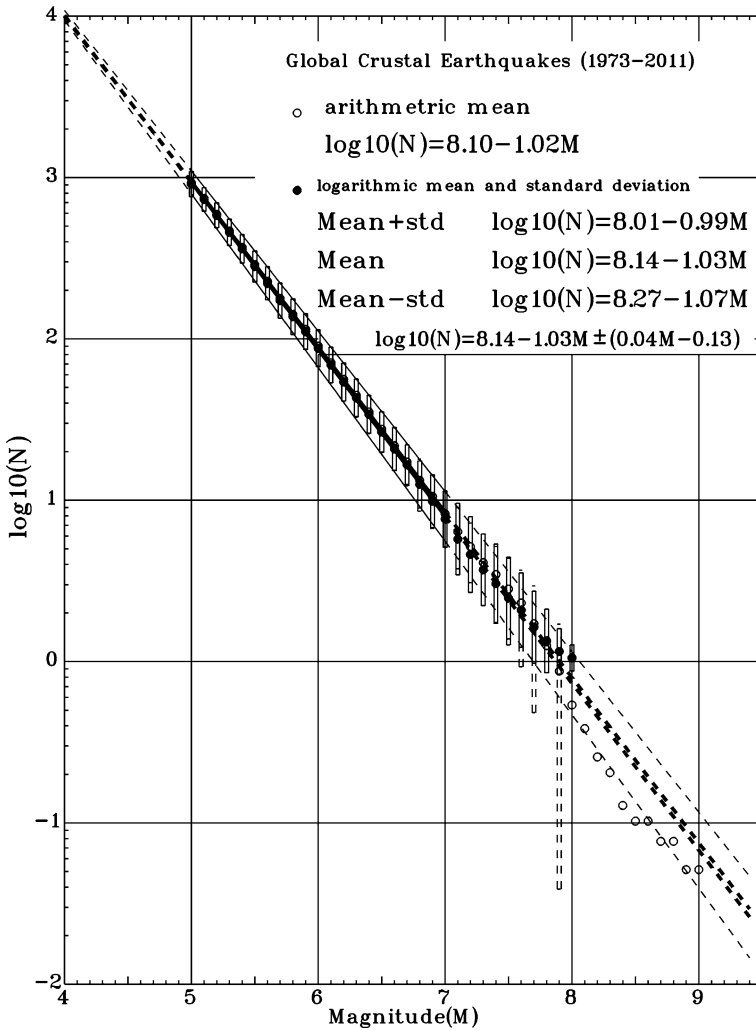


Fig. 8 Plots of the logarithmic and arithmetic mean annual seismicity rates of global crustal earthquakes with magnitudes $M \geq 5.0$. *Open circle* represents the observed arithmetic mean; *black dot* represents the observed logarithmic mean. *Solid bar* represents the *upper* and *lower* bounds of the logarithmic mean, and the *dashed bar* represents the *upper* and *lower* bounds of the arithmetic mean. The *solid lines* represent regression constrained by observed data, whereas *dashed lines* represent extrapolations

seismicity rate but also its dispersion for global crustal earthquakes with $M \geq 7.0$, as shown in Table 1.

The observed conventional arithmetic means are also shown for comparison by open circles with their corresponding regression line in Fig. 8. From the figure, we can see the observed arithmetic means are encompassed within the logarithmic mean \pm standard deviation range. The portion from $M5.0$ to $M7.8$ lies above and another portion from $M7.9$ to $M9.0$ below the logarithmic mean line. The deficiency in large events is probably due to the relative short time period of the NEIC catalog we selected for this study, so that not enough numbers of large earthquakes are included. This means that estimation on the

annual seismicity rates for global crustal earthquakes by Eq. 10 will cover the whole range of observed conventional arithmetic means. It should be pointed out further that Eq. 10 also shows an increasing logarithmic standard deviation with magnitude. This would imply increasing uncertainties with magnitude in the estimation of annual seismicity rates for large global crustal earthquakes.

5 Observed and estimated annual seismicity rates and their corresponding return periods for Taiwan and global crustal earthquakes

Finally, the analytical Eqs. 6 and 10 can be used to calculate the observed and estimated median annual seismicity rate and its dispersion at \pm standard deviation level and their corresponding return periods for Taiwan and global crustal earthquakes from $M5.0$ to $M9.0$, respectively. The results are listed in Table 1, with the observed values presented in bold-type numbers and the estimated values in regular-type numbers.

For Taiwan crustal earthquakes, Table 1 gives an observed median annual seismicity rate of 13.18 event/year with a range from 7.08 to 24.55 event/year for $M \geq 5.0$. The corresponding median return period is 0.076 years with a range from 0.041 to 0.14 years. The estimated median annual seismicity rate and its dispersion are 1.66 (0.85–3.24) event/year, 0.21 (0.10–0.43) event/year, 0.026 (0.012–0.056) event/year and 0.0033 (0.0015–0.0074) event/year for $M \geq 6.0$, $M \geq 7.0$, $M \geq 8.0$ and $M \geq 9.0$, respectively. The corresponding median return period and its dispersion are 0.60 (0.31–1.18) years, 4.76 (2.33–10.0) years, 38.46 (17.86–83.33) years and 303.03 (135.14–666.67) years for $M \geq 6.0$, $M \geq 7.0$, $M \geq 8.0$ and $M \geq 9.0$, respectively.

For global crustal earthquakes, Table 1 gives an observed median annual seismicity rate of 977.24 event/year with a range from 831.76 to 1148.16 event/year for $M \geq 5.0$. The corresponding median return period is 0.001 years with a range from 0.0009 to 0.0012 years. Additional observed median annual seismicity rate and its dispersion are 91.20 (70.79–117.49) event/year and 8.51 (6.03–12.02) event/year for $M \geq 6.0$ and $M \geq 7.0$, respectively. The estimated median annual seismicity rate and its dispersion are 0.79 (0.51–1.23) event/year and 0.074 (0.04–0.13) event/year for $M \geq 8.0$ and $M \geq 9.0$, respectively. The corresponding median return period and its dispersion are 0.011 (0.0085–0.014) years, 0.12 (0.083–0.17) years, 1.27 (0.75–1.96) years and 13.51 (7.69–25.00) years for $M \geq 6.0$, $M \geq 7.0$, $M \geq 8.0$ and $M \geq 9.0$, respectively. It is interesting to point out that the median annual seismicity rate of Taiwan crustal earthquakes with $M \geq 5.0$ accounts for about 1.35% of global crustal earthquakes.

6 Conclusions and discussion

The mean annual seismicity rate and its standard deviation are required for quantitative estimation of the probability of future earthquakes. Their applications in earthquake studies have advanced considerably in recent years. Relative occurrences between large and small earthquakes have been found to follow closely the Gutenberg–Richter (G–R) relation (Gutenberg and Richter 1941, 1944).

Conventionally, the G–R relation is represented in terms of the arithmetic mean annual seismicity rate, which is commonly obtained by dividing the total number of events with

the number of years of the catalog coverage. This conventional representation has the advantage of being straightforward. However, it has two major shortcomings. First, the arithmetic standard deviation cannot be incorporated explicitly in the log-linear G–R relation because of its asymmetry with respect to the arithmetic mean in logarithmic domain. Second, both the arithmetic mean and its standard deviation are susceptible to significant influence of spuriously large numbers of aftershocks from major earthquake sequences. These shortcomings are clearly illustrated by plotting the arithmetic mean and standard deviation in Fig. 7 for Taiwan crustal earthquakes from 1973 to 2011.

As an alternative we propose to represent the Gutenberg–Richter relation in terms of the logarithmic mean and its standard deviation, as given in Eq. 6 and shown in Fig. 7 for Taiwan crustal earthquakes, as well as in Eq. 10 and Fig. 8 for global crustal earthquakes. We can see in Fig. 7 both the logarithmic mean and its standard deviation are very well constrained in the magnitude range from $3.0 \leq M \leq 5.0$, where sufficiently large annual event numbers are available. Accordingly, the analytical Eq. 6 can be extrapolated to give robust estimates of the corresponding median annual seismicity rate as well as its upper and lower bounds at \pm one standard deviation for large crustal earthquakes with $M \geq 5.0$ in Taiwan, as shown in Table 1.

In order to further demonstrate the merits of our new method for greater magnitudes, we apply the same process to the global crustal earthquake data from the NEIC catalog. The results, as given in Eq. 10 and shown in Fig. 8, again show the Gutenberg–Richter relation is very well constrained in the magnitude range from $5.0 \leq M \leq 7.0$, where sufficiently large annual event numbers are available. The analytical Eq. 10 can be used to obtain robust estimates of the corresponding median annual seismicity rate as well as its lower and upper bounds at \pm one standard deviation for large global crustal earthquakes with $M \geq 7.0$, as shown in Table 1. It is interesting to see in the table that Taiwan crustal earthquakes with $M \geq 5.0$ account for about 1.35% of corresponding global crustal seismicity.

It is noted that the new approach would not be applicable if there are many zero or lower counts in earthquake occurrence in individual years. From the two data sets, we can see the new approach can yield a well-constrained Gutenberg–Richter relation if the logarithmic mean annual seismicity rate is greater than about 1.0, i.e., more than 10 events per year. Such a G–R relation can be extrapolated to obtain robust estimates on corresponding median annual seismicity rate and its dispersion for larger earthquakes.

In summary, our alternative representation of the Gutenberg–Richter relation provides a convenient analytical means to make robust assessment and estimation of both the median annual seismicity rate and its dispersion for any given magnitudes. Inclusion of the dispersion will account for the variability of individual annual seismicity rates, which is missing in the conventional representation in terms of only the mean annual seismicity rate. This alternative representation of the G–R relation can be used to improve the conventional probabilistic seismic hazard assessment.

7 Data and resources

Taiwan earthquake data are taken from published works listed in References. The global earthquake data are taken from the NEIC Web site: <http://earthquake.usgs.gov/earthquakes/search/>. Some plots were made using the Generic Mapping Tools version 4.3.1 (www.soest.hawaii.edu/gmt; Wessel and Smith 1998, last accessed August 2006).

Acknowledgements We appreciate greatly Dr. Thomas Glade, the Editor-In-Chief, Dr. Bhavani Sridhar, the Associate Editor, and two anonymous reviewers for their valuable comments and suggestions for improvements of this paper. We also wish to express special thanks to the staff of the Institute of Earth Sciences and the Central Weather Bureau for compiling the Taiwan earthquake catalog and to the staff of National Earthquake Information Center for compiling the global earthquake catalog. This study is supported by Ministry of Science and Technology (MOST).

References

- Aki K (1965) Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits. *Bull Earthq Res Inst* 43:237–239
- Chang WY, Chen KP, Tsai YB (2016) An updated and refined catalog of earthquakes in Taiwan (1900–2014) with homogenized M_w magnitudes. *Earth Planets Space* 68:45
- Chen KP, Tsai YB (2008) A catalog of Taiwan earthquakes (1900–2006) with homogenized M_w magnitudes. *Bull Seism Soc Am* 98:483–489
- Frankel A, Mueller C, Barnard T, Perkins D, Leyendecker EV, Dickman N, Hanson S, Hopper M (1996) National seismic-hazard maps: documentation, open-file report 96-532, U.S. Geological Survey, 71 p
- Gutenberg B, Richter CF (1941) Seismicity of the earth. *Geol Soc Am Special Papers* 34:1–131
- Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. *Bull Seism Soc Am* 34:185–188
- Gutenberg B, Richter CF (1954) Seismicity of the earth and associated phenomena. Princeton University Press, Princeton
- Hutton K, Woessner J, Hauksson E (2010) Earthquake monitoring in Southern California for seventy-seven years (1932–2008). *Bull Seism Soc Am* 100:423–446
- Konşuk H, Aktaş S (2013) Estimating the recurrence periods of earthquake data in Turkey. *Open J Earthq Res* 2:21–25
- Michael AJ (2014) How complete is the ISC-GEM global earthquake catalog? *Bull Seism Soc Am* 104:1829–1837
- Reasenber PA (1985) Second-order moment of central California seismicity, 1969–1982. *J Geophys Res* 90:5479–5495
- Reasenber PA, Jones LM (1989) Earthquake hazard after a mainshock in California. *Science* 265:1173–1176
- Reiter L (1990) Earthquake hazard analysis: issues and insights. Columbia University Press, New York, p 182
- Wessel P, Smith WHF (1998) New, improved version of generic mapping tools released. *EOS Trans AGU* 79:579