Journal of GeoEngineering, Vol. 16, No. 4, pp. 145-154, December 2021
http://dx.doi.org/10.6310/jog.202112_16(4).3

A NEW APPROACH FOR MODELING EARTHQUAKE MAGNITUDE PROBABILITY DISTRIBUTION

Min-Hao Wu [1], Jui-Pin Wang [2*], Jieh-Haur Chen [3], and Yan-Hui Lin [4]

ABSTRACT

For obtaining the (theoretical) earthquake magnitude probability distribution, the McGuire-Arabasz equation is commonly used with the *b*-value calibrated from the Gutenberg-Richter relationship. In this paper, we propose a new method, also based on the McGuire-Arabasz equation, while using the *b*-value directly calibrated from the observed magnitude probability distribution. We examined the new method with 149,879 earthquakes from five regions, and all of the tests show that the new approach outperforms the current method for modeling the observed magnitude probability distributions. More importantly, the new approach is as easy-to-use as the current method, with the optimization computation that can be easily completed by Excel Solver.

Key words: Earthquake magnitude probability distribution, optimization, statistical tests.

1. INTRODUCTION

In the past decades, earthquakes had caused a great deal of damage to our society, such as the 1999 Chi-Chi earthquake (Taiwan), the 2004 Sumatra-Andaman (Indonesia) earthquake, the 2008 Wenchuan (China) earthquake, and the 2011 Tohoku (Japan) earthquake. The main reason that we cannot prevent those earthquake-induced disasters is because the earthquake cannot be predicted (Geller *et al*. 1997), given that we cannot monitor the stress condition several kilometers below the ground surface where the focal points of earthquakes are located (Geller *et al*. 1997).

Seismic hazard analysis is for developing site-specific earthquake-resistant design parameters, based on the seismological evidence/data around the study site. One of the well-known seismic hazard assessments is PSHA (Cornell 1968), standing for probabilistic seismic hazard analysis. Nowadays, PSHA has been designated as the standard method for developing site-specific earthquake-resistant design ground motions for nuclear facilities (U.S. Nuclear Regulatory Commission 2007), not to mention that a number of PSHA case studies have also been reported in scientific journals (Cheng *et al.* 2007; Roshan and Basu 2010; Wang *et al.* 2013; Wang *et al.* 2016; Ayele 2017).

From the title of PSHA, it is understood that several parameters are treated as random variables in this probabilistic analysis, including earthquake magnitude. As a result, one of the input data of PSHA is the earthquake magnitude probability distribution (Fig. 1). Nowadays (to the best of our knowledge), the current method used for modeling the earthquake magnitude probability distributions was proposed by McGuire and Arabasz (1990), which was on the basis of the *b*-value from the Gutenberg-Richter recurrence law (Gutenberg and Richter 1944):

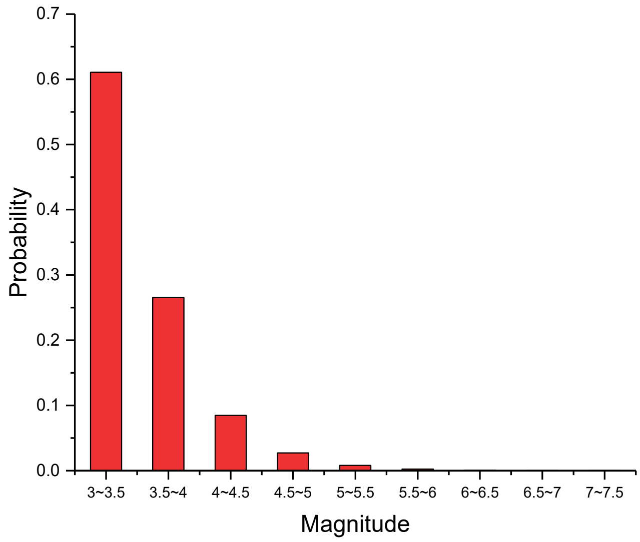


Fig. 1 Earthquake magnitude probability distribution

$$\log(N_{M\geq m}) = a - bm \quad (1)$$

where $N_{M \geq m}$ is the earthquake number with magnitude $M \geq m$. Note that m in this paper is referred to as magnitude of exceedance, or MOE. As a result, given that parameters a and b have been calibrated with seismicity data plotted in the space of $\log(N_{M \geq m})$ *vs.* $M \geq m$, we can estimate the earthquake number with magnitude $M \geq m$ as 10^{a-bm}.

As mentioned previously, McGuire and Arabasz (1990) utilized this relationship to develop the earthquake magnitude probability distribution, and the derivation is given in the following. Given the cutoff magnitude = m_0 and maximum magnitude = m_{max}, the number of earthquakes between m_0 and m_{max} is equal to:

Manuscript received June 15, 2021; revised October 1, 2021; accepted October 20, 2021.
[1] Professor, Dept. Civil & Environmental Engineering, National University of Kaohsiung, Taiwan.
[2*] Professor (corresponding author), Dept. Civil Engineering, National Central University; Research Center for Hazard Mitigation and Prevention, National Central University, Taiwan (e-mail: jpwang@ncu.edu.tw).
[3] Professor, Dept. Civil Engineering, National Central University, Taiwan.
[4] Master Student, Dept. Civil Engineering, National Central University, Taiwan.

$$N_{m_{\max} > M \geq m_0} = 10^{a - bm_0} - 10^{a - bm_{\max}} \tag{2}$$

Similarly, the number of earthquakes with magnitude between m_1 and m_2 (both between m_0 and $m_{\max}$) is:

$$N_{m_2 > M \geq m_1} = 10^{a - bm_1} - 10^{a - bm_2} \tag{3}$$

Then based on the fundamental of probability, the probability that the earthquake (magnitude) is between m_1 and m_2 given cutoff magnitude = m_0 and maximum magnitude = $m_{\max}$ becomes:

$$\Pr(m_1 \leq M < m_2; m_0, m_{\max}) = \frac{N_{m_1 \leq M < m_2}}{N_{m_0 \leq M < m_{\max}}} = \frac{10^{a - bm_1} - 10^{a - bm_2}}{10^{a - bm_0} - 10^{a - bm_{\max}}} \tag{4}$$

By dividing 10^a in the numerator and denominator, Eq. (4) becomes:

$$\Pr(m_1 \leq M < m_2; m_0, m_{\max}) = \frac{10^{-bm_1} - 10^{-bm_2}}{10^{-bm_0} - 10^{-bm_{\max}}} \tag{5}$$

Therefore, Eq. (5) is the McGuire-Arabasz approach for developing the earthquake magnitude probability distributions. To avoid confusion, we refer to the b-value in Eq. (5) as $b_{M\text{-}D}$ (standing for magnitude distribution) for developing the magnitude probability distributions, and refer to the b-value in Eq. (1) as $b_{G\text{-}R}$ (standing for Gutenberg Richter) for developing the Gutenberg-Richter recurrence law (linear regression line in the space of $\log N_{M \geq m}$ *vs.* $M \geq m$). However, it is noted that in the current practice, we use $b_{G\text{-}R} = b_{M\text{-}D}$ in the McGuire-Arabasz equation to develop the magnitude probability distribution, rather than re-calibrating $b_{M\text{-}D}$ based on the observed magnitude probability distribution directly. For example, Roshan and Basu (2010) conducted PSHA for a nuclear power plant site in India, and they followed the currently standard procedure using the b-value (or $b_{G\text{-}R}$) from the Gutenberg-Richter relationship to obtain the (theoretical) magnitude probability distribution.

The scope of this study is to introduce a new procedure for developing magnitude probability distributions. We used 149879 earthquake data to test the performance of the new approach. Based on the earthquake data from five different regions, all of the tests demonstrate the new approach outperforms the current method in modeling the observed magnitude probability distributions. More importantly, the new approach is as easy-to-use as the current method, with one optimization computation that can be easily completed by Excel Solver.

2. METHODOLOGY

2.1 USGS Database

The first task of this study is to collect earthquake data (or earthquake catalogs). In this study, we utilized the online database established by the U.S. Geological Survey (USGS) to collect earthquake catalogs. Figure 2 shows the interface of the USGS database (U.S. Geological Survey 2021). Basically, users simply need to input the cutoff magnitude, time frame, and the searching region, and then the system will compile the data of interest in an Excel file for the users to download. Another

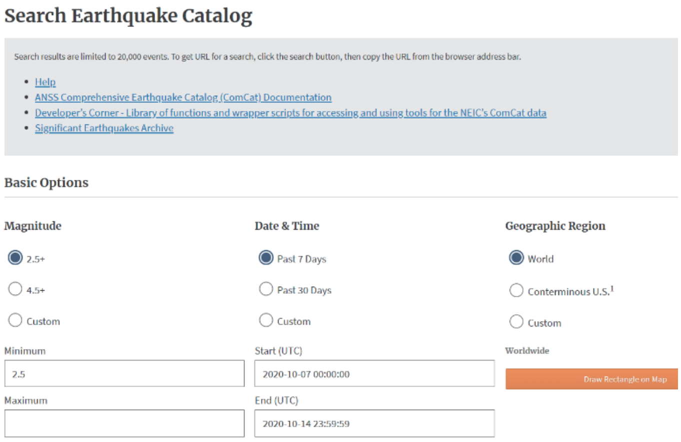


Fig. 2 The interface of the online USGS database (https://www.usgs.gov/natural-hazards/earthquake-hazards/earthquakes)

user-friendly function of the USGS database is that the users can "draw rectangle on map" (see the bottom right corner in Fig. 2) to circle the searching area of interest, instead of keying in the longitudes and latitudes of the searching area.

2.2 Magnitude Conversion

The data downloaded from the USGS database include the occurrence date, earthquake magnitude, depth, longitude, latitude, *etc.* However, it is noted that the unit of magnitude used in the database is inconsistent, including moment magnitude (M_w), local magnitude (M_L), surface wave magnitude (M_s), body wave magnitude (M_b), and duration/code magnitude (M_d). As a result, before we analyzed the data, we had to convert the different magnitudes into the same type. In this study, we converted them to moment magnitude because moment magnitude does not have the saturation issue (Kramer 1996). For converting them into M_w, the following relationships (Kadirioğlu and KarTal 2016) were used:

$$M_w = 1.209 M_b - 0.886 \tag{6}$$

$$M_w = 1.029 M_L + 0.227 \tag{7}$$

$$M_w = 1.951 M_d + 0.586 \tag{8}$$

$$M_w = 0.572 M_s + 2.249 \quad ; \quad M_s \leq 5.4 \tag{9}$$

$$M_w = 0.813 M_s + 1.172 \quad ; \quad M_s > 5.4 \tag{10}$$

2.3 Difference Between Observed and Simulated Probability Distributions

In this study, we will quantify the difference between the observed and simulated magnitude probability distributions (Fig. 3) for examining if the new approach can better simulate the observed distribution. We used the chi-square value (χ^2) to quantify the difference between model and observation, which can be expressed as follows (Ang and Tang 2007):

$$\chi^2 = \sum_{i=1}^{n} \frac{(t_i - o_i)^2}{t_i} \tag{11}$$

where t_i and o_i denote the theoretical and observational in the i-th data bin (Fig. 3). Understandably, the smaller the chi-square value is, the better fitting between model and observation will be, and when the chi-square value = 0, the model and observation are identical or perfectly matched.

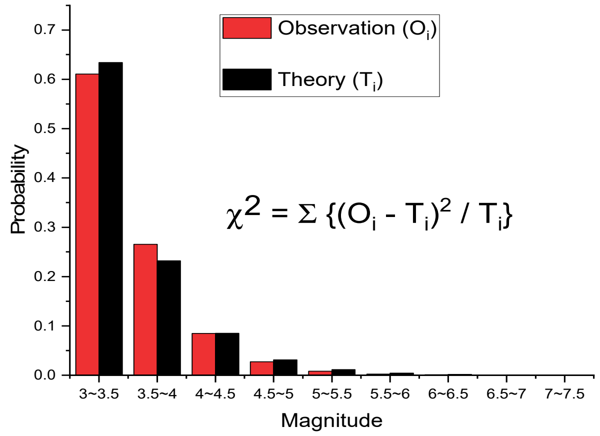


Fig. 3 **Using the chi-square value to quantify the difference between model and observation in this study**

2.4 Calculating the Optimal b_{M-D}

As mentioned previously, the scope of this study is to search for the optimal b_{M-D} that can best capture the observed magnitude probability distribution using the McGuire-Arabasz equation (Eq. (5)). Expectedly, the optimization calculation could be tedious. However, such an optimization can be easily solved with Excel Solver or other computer codes alike. In this study, we used Excel Solver for those optimization computations to find the optimal b_{M-D}.

3. RESULTS

3.1 Taiwan

The first test is based on the earthquake data around Taiwan. Figure 4(a) shows the region around Taiwan within which 3,215 earthquakes from 1980/01/01 to 2020/09/30 were collected from the USGS database. Note that the longitudes and latitudes of the searching region are shown in the bottom right corner of Fig. 4(a).

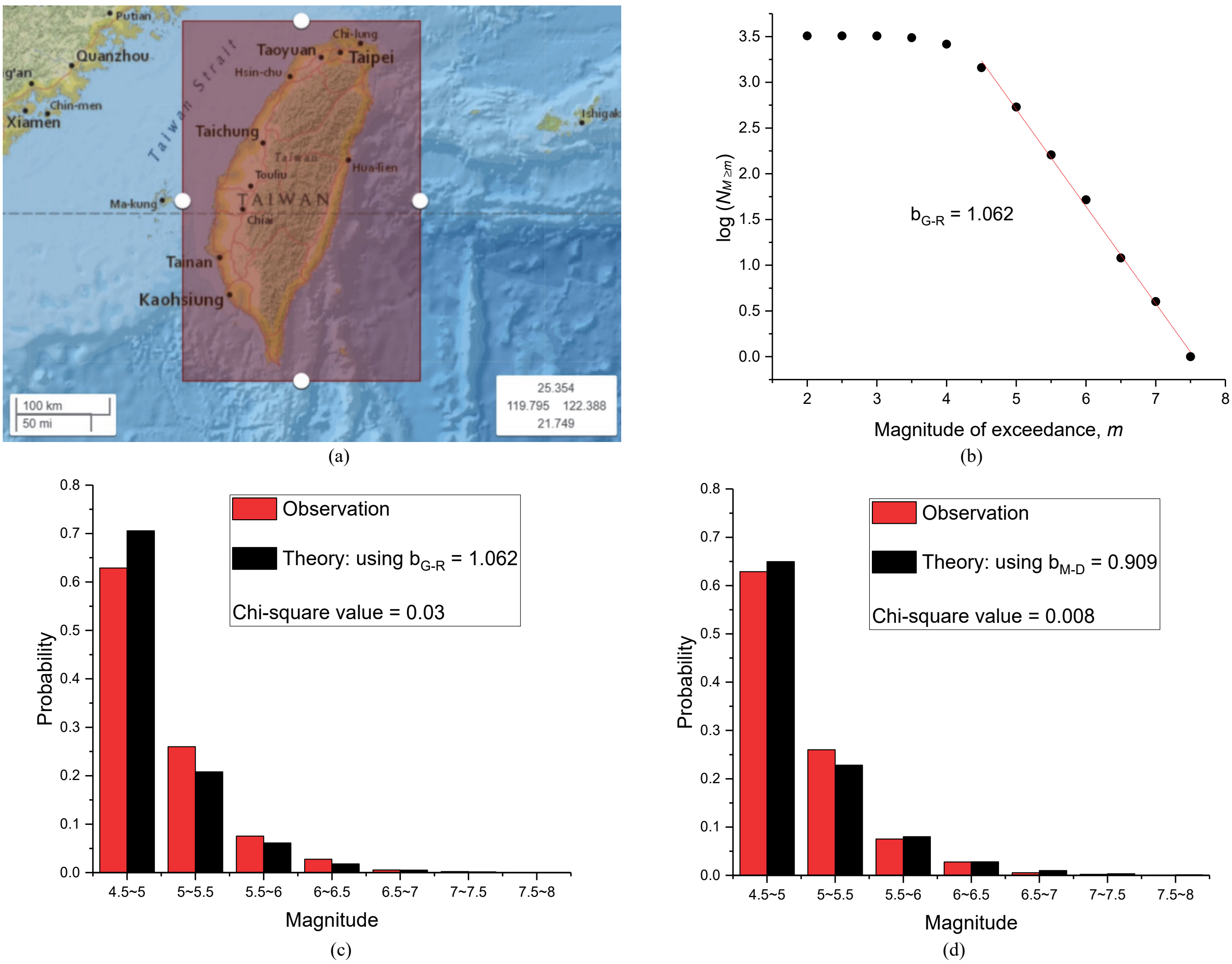


Fig. 4 **The results for the "Taiwan" case: (a) The searching area around Taiwan; (b) the G-R relationship; (c) observed magnitude probability distribution and the simulation using the current approach; (d) observed magnitude probability distribution and the simulation using the new approach**

Figure 4(b) shows the G-R (Gutenberg-Richter) relationship of the earthquake data. Based on the suggestion of Rydelek and Sacks (1989), the point starts separating from the straight line fit with the larger-magnitude data points implying that below this magnitude point, the earthquake data should be incomplete. As a result, the earthquakes above M_w 4.5 should be complete in this case (the point of M_w 4.0 starts separating from the straight line). Based on the "complete" data (magnitude above M_w 4.5), the value of $b_{G\text{-}R}$ is equal to 1.062.

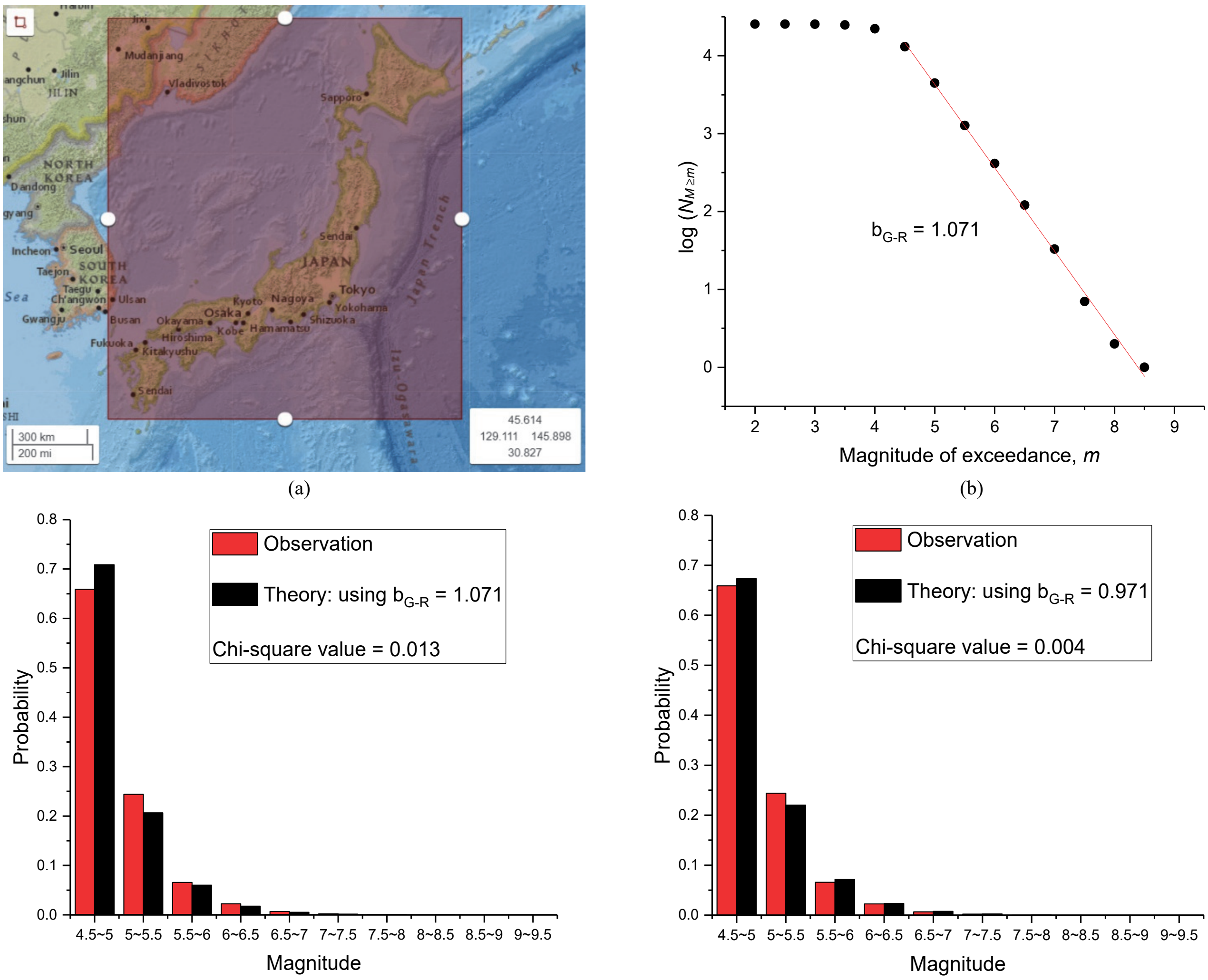
Figure 4(c) shows the observed magnitude probability distribution based on the same pool of data, as well as the simulation developed with the current method by using $b_{G\text{-}R} = b_{M\text{-}D} = 1.062$ in the McGuire-Arabasz equation. It shows that the fitting is satisfactory, and the difference in terms of the chi-square value is equal to 0.03.

Figure 4(d) shows the same observed magnitude probability distribution and the simulation developed with new method using the optimal $b_{M\text{-}D} = 0.909$. The chi-square value between the two was lowered to 0.008. Comparing Fig. 4(c) with Fig. 4(d), we found that the overestimation issue in low-magnitude ranges present in the current practice was improved by the new approach. In terms of the chi-square value, the new simulation (using optimal $b_{M\text{-}D}$) with the chi-square value of 0.008 is better than the current practice using $b_{G\text{-}R} = b_{M\text{-}D}$ with a chi-square value of 0.03.

3.2 Japan

The second test is based on earthquake data around Japan. Figure 5(a) shows the searching region with the same time frame from 1980/01/01 to 2020/09/30. A total of 25509 earthquakes were collected accordingly. Figure 5(b) shows the G-R relationship of the 25509 earthquakes, and the magnitude of completeness should be equal to M_w 4.5. Based on the "complete" data (magnitude above M_w 4.5), the value of $b_{G\text{-}R}$ is equal to 1.071.

Figure 5(c) shows the observed magnitude probability distribution, and the simulation of the current method using $b_{G\text{-}R} = b_{M\text{-}D} = 1.071$ in the McGuire-Arabasz equation. The difference (chi-square value) between the two is equal to 0.013. Figure 5(d) shows the same observed magnitude probability distribution and the simulation from the new method using the optimal $b_{M\text{-}D} = 0.971$. Given that a lower chi-square value of 0.004 (compared to 0.013) was obtained, it also implies that the new approach using the optimal $b_{M\text{-}D}$ outperforms the current practice using $b_{G\text{-}R} = b_{M\text{-}D}$ in the McGuire-Arabasz equation to develop theoretical earthquake magnitude probability distributions.



(a) (b) (c) (d)

Fig. 5 The results for the "Japan" case: (a) The searching area around Japan; (b) the G-R relationship; (c) observed magnitude probability distribution and the simulation using the current approach; (d) observed magnitude probability distribution and the simulation using the new approach

3.3 California

The third test is based on 70898 earthquakes around California collected from the USGS database. Figure 6(a) shows the searching region within the same time frame (1980/01/01 to 2020/09/30). Figure 6(b) shows the G-R relationship of the 70898 earthquake data, and the magnitude of completeness should be equal to M_w 3.0. Based on the "complete" data (magnitude above M_w 3.0), the value of $b_{G\text{-}R}$ is equal to 0.924.

Figure 6(c) shows the observed magnitude probability distribution and the simulation of the current practice using $b_{G\text{-}R} = b_{M\text{-}D} = 0.924$ in the McGuire-Arabasz equation. The difference between the model and observation in terms of chi-square value is equal to 0.011. Figure 6(d) shows the same observed magnitude probability distribution along with the new simulation using the optimal $b_{M\text{-}D} = 0.873$ in the McGuire-Arabasz equation, which leads to a lower chi-square value equal to 0.008 (compared to 0.011).

3.4 Iran

The fourth test is based on 5058 earthquakes around Iran that were collected from the USGS database. Figure 7(a) shows the searching region around Iran within the same time frame from 1980/01/01 to 2020/09/30. Figure 7(b) shows the G-R relationship, and the magnitude of completeness should be equal to M_w 4.5. Based on the "complete" data, the value of $b_{G\text{-}R}$ is equal to 1.205.

Figure 7(c) shows the observed magnitude probability distribution and the simulation of the current practice using $b_{G\text{-}R} = b_{M\text{-}D} = 1.205$ in the McGuire-Arabasz equation. The difference in terms of chi-square value is equal to 0.03. Figure 7(d) shows the observed magnitude probability distribution and the new simulation using the optimal $b_{M\text{-}D} = 1.051$ in the McGuire-Arabasz equation, leading to a lower chi-square value of 0.014 (compared to 0.03).

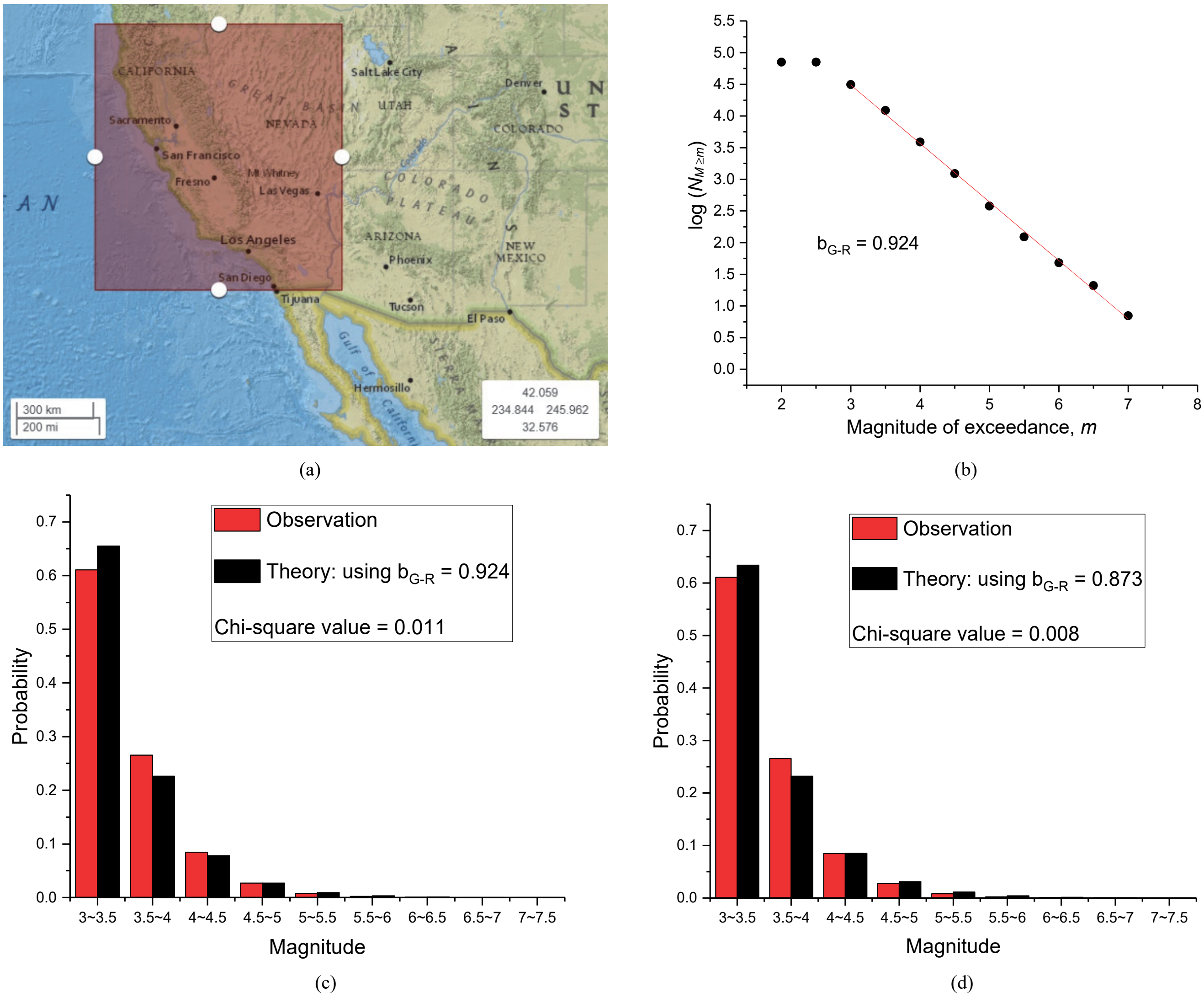


Fig. 6 The results for the "California" case: (a) The searching area around California; (b) the G-R relationship; (c) observed magnitude probability distribution and the simulation using the current approach; (d) observed magnitude probability distribution and the simulation using the new approach

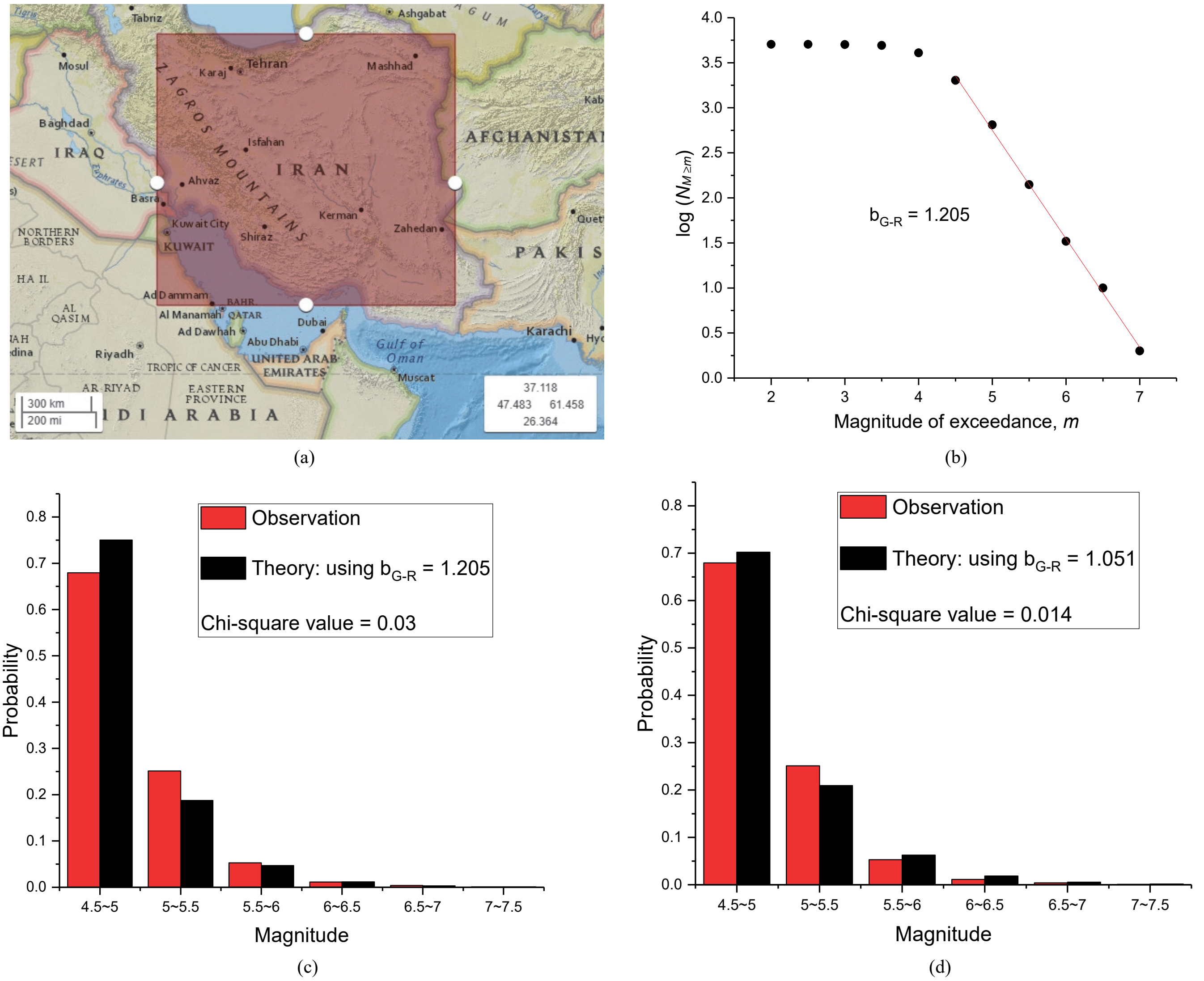


Fig. 7 The results for the "Iran" case: (a) The searching area around Iran; (b) the G-R relationship; (c) observed magnitude probability distribution and the simulation using the current approach; (d) observed magnitude probability distribution and the simulation using the new approach

3.5 Turkey and Greece

The fifth test is based on 45199 earthquakes around Turkey and Greece collected from the USGS database. Figure 8(a) shows the searching region within the same time frame from 1980/01/01 to 2020/09/30. Figure 8(b) shows the G-R relationship, and the magnitude of completeness should be close to M_w 3.5. Based on the "complete" data, the value of $b_{G\text{-}R}$ is equal to 1.086.

Figure 8(c) shows the observed magnitude probability distribution and the model based on the current practice using $b_{G\text{-}R} = b_{M\text{-}D} = 1.086$ in the McGuire-Arabasz equation. The difference (chi-square value) is equal to 0.002. Figure 8(d) shows the observed magnitude probability distribution and the new simulation using the optimal $b_{M\text{-}D} = 1.068$, also leading to a lower chi-square value of 0.001 (compared to 0.002).

4. DISCUSSIONS

4.1 Empirical Relationship between $b_{G\text{-}R}$ and $b_{M\text{-}D}$

Table 1 summarizes the values of $b_{G\text{-}R}$, $b_{M\text{-}D}$, and chi-square values of the tests. Expectedly, using the optimal $b_{M\text{-}D}$ in the McGuire-Arabasz equation (Eq. (5)) to model the magnitude probability distribution outperformed the simulation of the current practice using $b_{G\text{-}R} = b_{M\text{-}D}$. This is because $b_{G\text{-}R}$ is not the optimal value for fitting the observation in the space of magnitude probability *vs.* magnitude (*e.g.,* Fig. 4(c)). Instead, $b_{G\text{-}R}$ is the optimal value for fitting the data in the space of $\log N_{M \geq m}$ *vs.* $M \geq m$ (*e.g.,* Fig. 4(b)).

Table 1 Summary of using different values of $b_{M\text{-}D}$ to fit the observed magnitude probability distribution

Region	$b_{M\text{-}D} = b_{G\text{-}R}$	χ^2	Optimal $b_{M\text{-}D}$	χ^2	$b_{M\text{-}D}$ by Eq. (12)	χ^2
Taiwan	1.062	0.030	0.909	0.008	0.969	0.011
Japan	1.071	0.013	0.971	0.004	0.975	0.004
California	0.924	0.011	0.873	0.008	0.875	0.008
Iran	1.205	0.030	1.051	0.014	1.067	0.014
Greece and Turkey	1.086	0.002	1.068	0.001	0.986	0.002

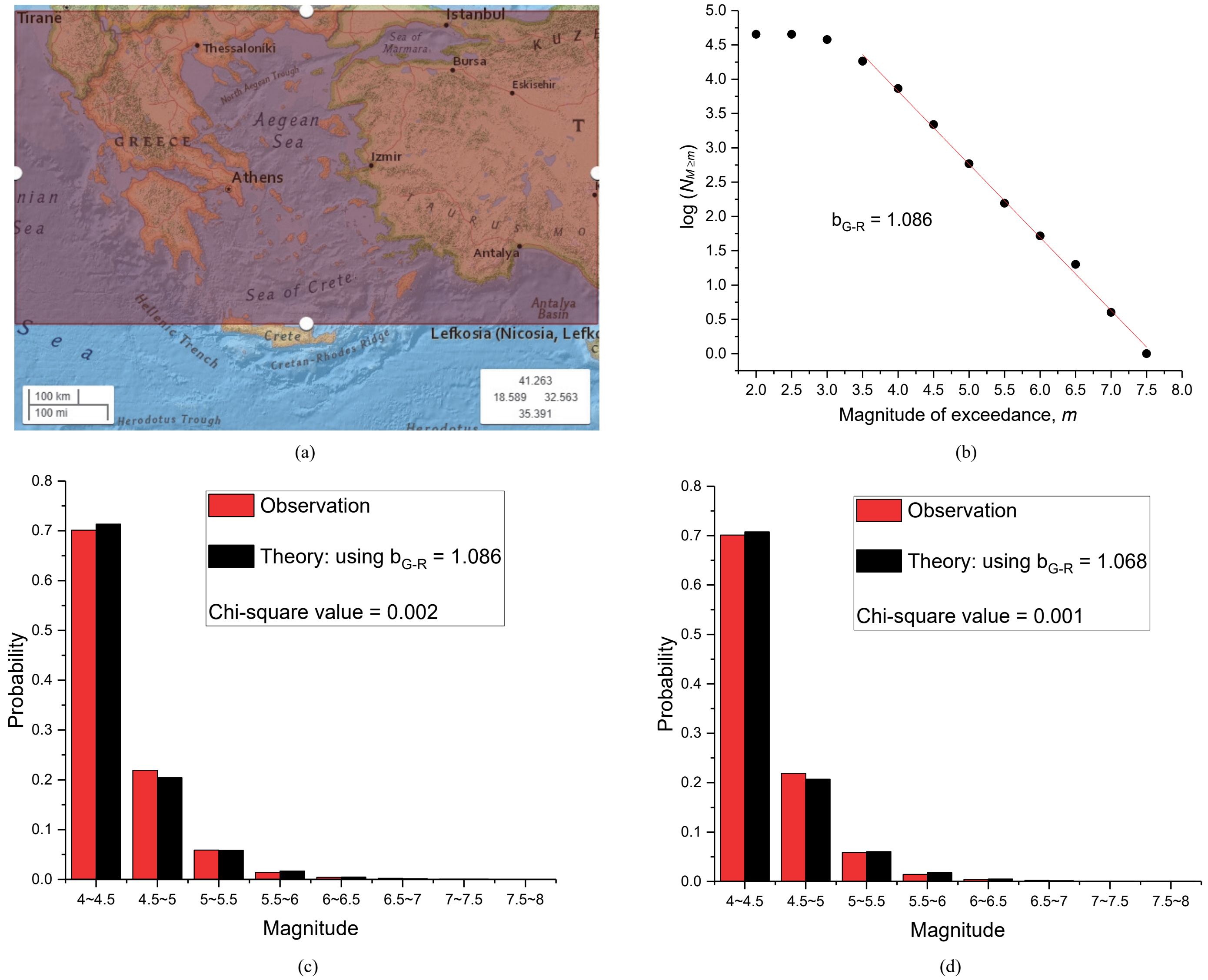


Fig. 8 The results for the "Greece and Turkey" case: (a) The searching area around Greece and Turkey; (b) the G-R relationship; (c) observed magnitude probability distribution and the simulation using the current approach; (d) observed magnitude probability distribution and the simulation using the new approach

It is also found that all of the tests show that $b_{G\text{-}R}$ is larger than the optimal $b_{M\text{-}D}$. Figure 9 shows the simple linear regression between the two from the five tests:

$$b_{G-R} = 0.683 b_{M-D} + 0.244 \tag{12}$$

Note that the coefficient of determination (R^2) of this regression is 0.68. Table 1 also summarizes the $b_{M\text{-}D}$ obtained with this empirical relationship. It shows that using the $b_{M\text{-}D}$ values from the empirical relationship to develop the magnitude probability distributions can also outperform (with a lower chi-square value) the current practice using $b_{G\text{-}R} = b_{M\text{-}D}$. As a result, if one finds it difficult to conduct the optimization computation to obtain the optimal $b_{M\text{-}D}$, another easier alternative is to utilize the empirical relationship to find $b_{M\text{-}D}$ from $b_{G\text{-}R}$, and use the $b_{M\text{-}D}$ in the McGuire-Arabasz equation.

One could comment that the current practice (using $b_{G\text{-}R} = b_{M\text{-}D}$) already achieved a satisfactory modeling of the earthquake magnitude probability distribution, why bother using the new approach? The simple response is that the new approach using

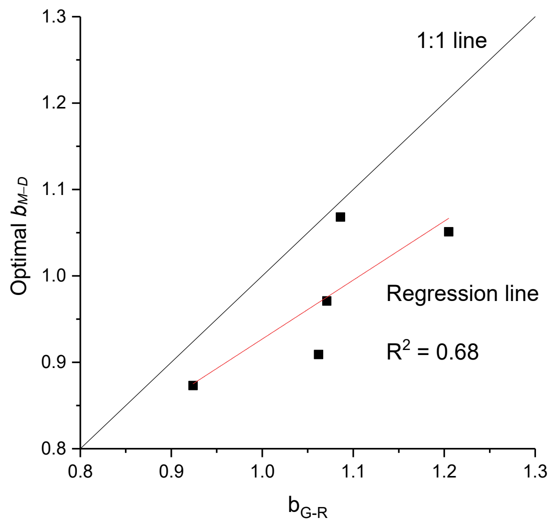


Fig. 9 The linear regression between $b_{G\text{-}R}$ and the optimal $b_{M\text{-}D}$ based on earthquake data from the five regions

the optimal b_{M-D} in the McGuire-Arabasz equation can outperform the current practice in the modeling of magnitude probability distributions. Furthermore, the new approach is as easy-to-use as the current practice, simply needing to conduct one optimization computation that can be conducted with Excel Solver or other computer codes alike.

4.2 Effect of Earthquake Catalog on the Proposed Method

This study used the USGS catalog to demonstrate the new method is better than the current method in the modeling of the observed magnitude probability distribution, which is attributed to the optimal b_{M-D} is obtained from the data plotted in the space of probability (P) *vs.* magnitude (M), which is the same data space as the magnitude probability distribution; by contrast, b_{G-R} is obtained from the data plotted in the space of $\log N_{M \geq m}$ *vs.* $M \geq m$, which is a different data space from that used for plotting the magnitude probability distribution (P *vs.* M).

Without doubt, USGS is a highly regarded institute in geological investigation and research, and its publications, such as earthquake catalogs, are widely recognized and accepted by the industry and academia. As a matter of fact, like this study, other researchers also used the USGS catalog (as it is) to examine and demonstrate their theories. For instance, Geng *et al*. (2021) utilized the USGS catalog to demonstrate the so-called nonparametric Bayesian method to estimate the spatial distribution of earthquake intensities; similarly, Wu *et al.* (2019) used the catalog to prove that earthquake magnitude should be a random variable following the Weibull distribution. Other studies include a PSHA project using the USGS catalog to establish the seismic hazard maps for the Tennessee areas (Wheeler and Mueller 2001), as well as the research based on the USGS catalog to characterize the earthquake potential along the Himalayan arc (Sharma *et al.* 2020). More importantly, among the aforementioned studies, the researchers did not point out any probable flaws with the USGS catalog.

Nevertheless, to further demonstrate our theory, we used another earthquake catalog published by another highly regarded research institute, the International Seismological Centre (http://www.isc.ac.uk/iscgem/download.php). However, unlike the USGS database that allows users to search for earthquake records for the area of interest, the ISC catalog lists worldwide earthquakes in one single file.

Figure 10 shows the comparison between the new and conventional methods to simulate the magnitude probability distribution, on the basis of the ISC catalog containing the worldwide earthquakes above M_w 5.5 from 2010 to 2016. As expected, the analysis verified the finding, once again, that using the optimal b-value calibrated with the data plotted in the space of P *vs.* M can better simulate the magnitude probability distribution (also in the data space of P *vs.* M), than using the b-value calibrated with the data plotted in a different data space like $\log N_{M \geq m}$ *vs.* $M \geq m$. As a result, this is the fundamental cause that the proposed optimization method is superior to the conventional methods for simulating the magnitude probability distribution, simply because the optimal b-value was obtained from the same data space (P *vs.* M) as the magnitude probability distribution.

As demonstrated with the USGS and ISC earthquake catalogs, the new method for simulating the magnitude probability distribution can be used for any earthquake catalog, while the same outcome must be obtained: the new optimization method is superior to the conventional methods for modeling the magnitude probability distribution owing to the nature of optimization. It is beyond the scope of this study to discuss the potential flaws of an earthquake catalog, despite that it was published by reputable institutes like USGS and ISC that specialize in geological and seismological investigation and research.

4.3 Effect of Maximum Earthquake Magnitude on the Proposed Method

Several approaches had been proposed for estimating the maximum earthquake magnitude ($m_{\max}$) based on instrumental earthquake records within a limited time span. For instance, Kijko (2004) proposed a non-parametric procedure to estimate the $m_{\max}$ for a region of interest, which does not need to assume what probability distribution the earthquake magnitude should be following. On the other hand, Pisarenko *et al.* (1996) proposed an estimator for $m_{\max}$ based on the fundamentals of probability and statistics, which was proved unbiased and associated with the lowest variance.

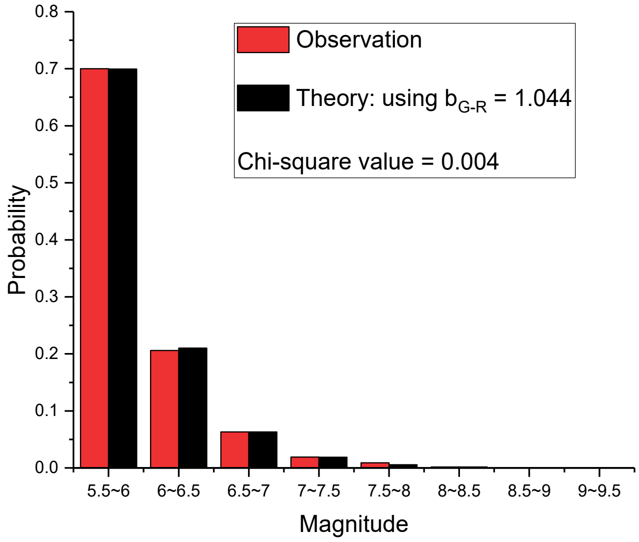
However, as mentioned in Sections 4.1 and 4.2, the scope of this study is to introduce a new method that was proved more superior to the conventional method for simulating the magnitude probability density, which is because the optimal b-value proposed by this study is calibrated with the data plotted in the same data space (*i.e., P vs. M*), so it can fit the observed magnitude probability distribution better than other methods using the b-value calibrated with data plotted in a different data space (*i.e.,* $\log N_{M \geq m}$ *vs.* $M \geq m$).

Previously, we have shown (Fig. 10) that the proposed method is better than the conventional methods no matter which earthquake catalogs (USGS or ISC) were tested. Likewise, the situation is applicable to $m_{\max}$. That is, regardless of the $m_{\max}$ value used for determining the magnitude probability distribution and b-value, using the optimal b-value calibrated with data plotted in the same data space as the magnitude probability distribution can better fit the observed magnitude probability dsitribution than using other b-values calibrated with data plotted in different data spaces. Therefore, no matter which $m_{\max}$ value was used for constructing the observed magnitude probability distribution, the proposed method can better model it because the parameter is calibrated with the data plotted in the same data space.

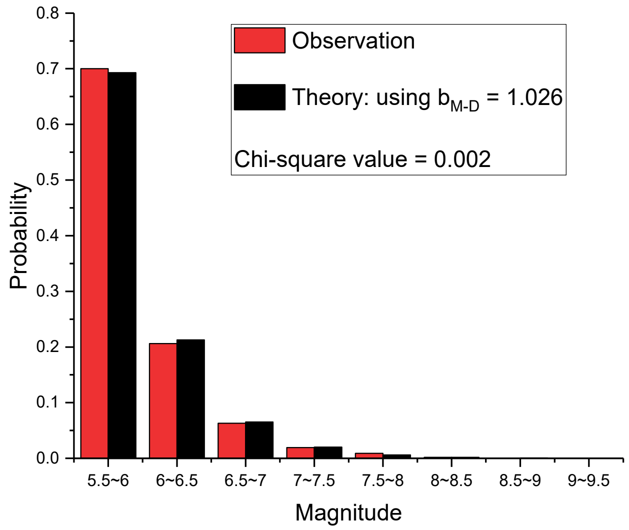
4.4 The "Delog" Effect

It is puzzling that the currently standard method for developing magnitude probability function is imprecise sometimes, and we attributed it to the "delog" effect, because the G-R law uses $\log N$ (earthquake number) in Y axis, and magnitude probability distribution uses N in Y axis.

We used the following hypothetical example to illustrate the substantial change from $\log N$ to N after "delog." Given $N = 550$ for M between 4 to 4.5, $N = 90$ for M between 4.5 to 5, $N = 15$ for M between 5 to 5.5, and $N = 1$ for M between 5.5 to 6; Accordingly, the G-R regression model has an R^2 as high as 0.99. However, based on the regression model, the predicted $\log(N)$ with $M \geq 4$ is equal to 2.902, so N is equal to 799 after "delog." And what is the observed N with $M \geq 4$? It is only 656. Between 656 (observation) and 799 (theory), the difference can reach 20% due to the "delog" effect, even though the difference between the observed $\log N$ (= 2.816) and expected $\log N$ (= 2.902) is "deceivingly" small.



(a) Observed magnitude probability distribution and the simulation using the current approach



(b) Observed magnitude probability distribution and the simulation using the new approach

Fig. 10 The test based on the ISC earthquake catalog

5. CONCLUSIONS

The earthquake magnitude probability distribution is one of the essential input data for PSHA. This paper proposed a new approach to improve the simulation of earthquake magnitude probability distributions. Based on the tests on the earthquake data from five regions, the following conclusions were drawn:

1. We tested the new approach with earthquake data from five different regions. All of the tests show that the new approach using the optimal $b_{M\text{-}D}$ in the McGuire-Arabasz equation outperforms the current practice using $b_{G\text{-}R} = b_{M\text{-}D}$ in the same equation.
2. The new approach is also as easy-to-use as the current practice, simply needing to conduct one optimization that can be easily conducted with Excel Solver.
3. An empirical relationship between $b_{G\text{-}R}$ and $b_{M\text{-}D}$ is also provided in this study. It shows that using the $b_{M\text{-}D}$ values from the empirical relationship also outperforms the current practice (using $b_{G\text{-}R} = b_{M\text{-}D}$), an even simpler option to improve the simulation of earthquake magnitude probability distributions.
4. Due to the power of optimization, the optimization-based method must provide the best-fit magnitude probability function compared to other methods. This will not be affected no matter which earthquake catalog, maximum earthquake magnitude, *etc.* are used.

ACKNOWLEDGEMENTS

The comments from three anonymous reviewers are highly appreciated.

FUNDING

We are also thankful for the financial support from the Ministry of Science and Technology of Taiwan on this research (Project No. 107-2221-E-008-021-MY3).

DATA AVAILABILITY

All data and/or computer codes used/generated in this study are included in this paper.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

REFERENCES

Ayele, A. (2017). "Probabilistic seismic hazard analysis (PSHA) for Ethiopia and the neighboring region." *Journal of African Earth Sciences*, **134**, 257-264. https://doi.org/10.1016/j.jafrearsci.2017.06.016

Ang, A.H.S. and Tang, W.H. (2007). *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering*. 2nd Ed., John Wiley & Sons. ISBN: 978-0-471-72064-5.

Cornell, C.A. (1968). "Engineering seismic risk analysis." *Bulletin of the Seismological Society of America*, **58**(5), 1583-1606. https://doi.org/10.1785/BSSA0580051583

Cheng, C.T., Chiou, S.J., Lee, C.T., and Tsai, T.B. (2007). "Study on probabilistic seismic hazard maps of Taiwan after Chi-Chi earthquake." *Journal of GeoEngineering*, **2**(1), 19-28. http://dx.doi.org/10.6310/jog.2007.2(1).3

Geller, R.J., Jackson, D.D., Kagan, Y.Y., and Mulargia, F. (1997). "Earthquakes cannot be predicted." *Science*, **275**, 1616-1617. https://doi.org/10.1126/SCIENCE.275.5306.1616

Gutenberg, B. and Richter, C.F. (1944). "Frequency of earthquake in California." *Bulletin of the Seismological Society of America*, **34**, 185-188. https://doi.org/10.1785/BSSA0340040185

Geng, J., Shi, W., and Hu, G. (2021). "Bayesian nonparametric nonhomogeneous Poisson process with application to USGS earthquake data." *Spatial Statistics*, **41**, 100495. https://doi.org/10.1016/j.spasta.2021.100495

International Seismological Centre (ISC). *The ISC-GEM Global Instrumental Earthquake Catalog*. Retrieved from http://www.isc.ac.uk/iscgem/download.php. Accessed on May 1, 2021.

Kramer, S.L. (1996). *Geotechnical Earthquake Engineering*. Prentice Hall Inc. ISBN: 0-13-374943-6.

Kadirioğlu, F.T. and Kartal, R.F. (2016). "The new empirical magnitude conversion relations using an improved earthquake catalog for Turkey and its near vicinity (1900-2012)." *Turkish Journal of Earth Sciences*, **25**(4), 300-310. https://doi.org/10.3906/yer-1511-7

Kijko, A. (2004). "Estimation of the maximum earthquake magnitude, m_{max}." *Pure and Applied Geophysics*, **161**, 1655-1681. https://doi.org/10.1007/s00024-004-2531-4

McGuire, R.K. and Arabasz, W.J. (1990). "An introduction to probabilistic seismic hazard analysis." *Geotechnical and Environmental Geophysics*, **1**, 333-353. https://doi.org/10.1190/1.9781560802785.ch12

Rydelek, P. and Sacks, I. (1989). "Testing the completeness of earthquake catalogues and the hypothesis of self-similarity." *Nature*, **337**, 251–253. https://doi.org/10.1038/337251a0

Roshan, A.D. and Basu, P.C. (2010). "Application of PSHA in low seismic region: A case study on NPP site in peninsular India." *Nuclear Engineering and Design*, **240**(10), 3443-3454. https://doi.org/10.1016/j.nucengdes.2010.04.037

Sharma, Y., Pasari, S., Ching, K.E., Dikshit, O., Kato, T., Malik, J.N., Chang, C.P., and Yen, J.Y. (2020). "Spatial distribution of earthquake potential along the Himalayan arc." *Tectonophysics*, **791**, 228556. https://doi.org/10.1016/j.tecto.2020.228556

U.S. Nuclear Regulatory Commission. (2007). *A Performance-Based Approach to Define the Site-Specific Earthquake Ground Motion*. Regulatory Guide 1.208. USNRC, Washington, DC.

U.S. Geological Survey (USGS). *Search Earthquake Catalog*. https://earthquake.usgs.gov/earthquakes/search/. Accessed on May 1, 2021.

Pisarenko, V.F., Lyubushin, A.A., Lysenko, V.B., and Golubieva, T.V. (1996). "Statistical Estimation of Seismic Hazard Parameters: Maximum Possible Magnitude and Related Parameters." *Bulletin of the Seismological Society of America*, **86**, 691-700. https://doi.org/10.1016/j.cageo.2012.09.021

Wang, Y.J., Chan, C.H., Lee, Y.T., Ma, K.F., Shyu, J.B.H., Rau, R.J., and Cheng, C.T. (2016). "Probabilistic seismic hazard assessment for Taiwan." *Terrestrial, Atmospheric and Oceanic Sciences*, **27**, 325-340. https://doi.org/10.3319/TAO.2016.05.03.01(TEM)

Wu, M.H., Wang, J.P., and Ku, K.W. (2019). "Earthquake, Poisson and Weibull distributions." *Physica A: Statistical Mechanics and Its Applications*, **526**, 121001. https://doi.org/10.1016/j.physa.2019.04.237

Wheeler, R.L. and Mueller, C.S. (2001). "Central US earthquake catalog for hazard maps of Memphis, Tennessee." *Engineering Geology*, **62**, 19-29. https://doi.org/10.1016/S0013-7952(01)00046-1